

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

tein search, using sw model

April 23, 2004, 14:00:14 ; Search time 59 Seconds
(without alignments)
857.220 Million cell updates/sec

US-09-147-801d-4

1 KOLMEIARMPVYKGBTEP.....QVRMGNDKFMVPTTSNIW 179

OLIGO

Gapop 60.0 , Gapext 60.0

1586107 seqs, 282547505 residues

0

hits satisfying chosen parameters: 1586107

ength: 0

ength: 2000000000

Listing first 45 summaries

A Geneseq 29Jan04: *
1: Geneseqp1980s: *
2: Geneseqp1990s: *
3: Geneseqp2000s: *
4: Geneseqp2001s: *
5: Geneseqp2002s: *
6: Geneseqp2003as: *
7: Geneseqp2003bs: *
3: Geneseqp2004s: *

is the number of results predicted by chance to have a
per than or equal to the score of the result being printed,
lived by analysis of the total score distribution.

SUMMARIES

\$	Query	Match	Length	DB	ID	Description
00.0	179	2	AAW56274		AAW56274 Picornavi	
3.9	23	2	AAY36668		Aay36668 Fragment	
3.9	23	6	ADAl1922		ADal1922 Human nov	
3.9	26	2	AAY36649		Aay36649 Fragment	
3.9	26	6	ADAl1903		Adal1903 Human nov	
3.9	72	4	AAM37746		Aam37746 Peptide #	
3.9	72	4	AAW77564		Aam77564 Human bon	
3.9	72	4	AAW64808		Aam64808 Human bra	
3.9	72	4	ABG59207		ABG59207 Human liv	
3.9	72	5	ABG46592		ABG46592 Human pep	
3.9	88	4	AAW15491		Aam15491 Peptide #	
3.9	88	4	AAW27979		Aam27979 Peptide #	
3.9	88	4	ABG29325		ABG29325 Peptide #	
3.9	88	4	ABBl9908		ABb19908 Protein #	
3.9	88	4	AAW67682		Aam67682 Human bon	
3.9	88	4	ABG49316		ABG49316 Human liv	
3.9	88	4	AAW03248		Aam03248 Peptide #	
3.9	88	5	ABG37251		ABg37251 Human pep	
3.9	93	4	AAW17078		Aam17078 Peptide #	
3.9	93	4	ABB36080		ABb36080 Peptide #	
3.9	93	4	AAW29572		Aam29572 Peptide #	
3.9	93	4	ABB30898		ABb30898 Peptide #	
3.9	93	4	ABB21472		ABb21472 Protein #	
3.9	93	4	AAW69249		Aam69249 Human bon	
3.9	93	4	AAW56862		Aam56862 Human bra	

26 7 3.9 93 4 AAM04779
27 7 3.9 93 5 ABG38859
28 7 3.9 135 2 AAY36655
29 7 3.9 135 6 ADAl1909
30 7 3.9 223 2 AAW71593
31 7 3.9 252 5 ABP70131
32 7 3.9 266 5 ABP69056
33 7 3.9 290 5 AAO15814
34 7 3.9 319 2 AAW78917
35 7 3.9 319 5 AAO15807
36 7 3.9 331 2 AAW71592
37 7 3.9 334 2 AAW99661
38 7 3.9 334 6 ADAl1633
39 7 3.9 341 5 ABP70129
40 7 3.9 350 5 AAO15812
41 7 3.9 357 5 AAO15809
42 7 3.9 361 3 AAB42023
43 7 3.9 389 5 ABB90814
44 7 3.9 444 4 AAU33986
45 7 3.9 460 4 ABB60884

ALIGNMENTS

RESULT 1
AAW56274
ID AAW56274 standard; protein; 179 AA.
XX AC AAW56274;
XX DT 27-AUG-2003 (revised)
XX DT 30-JUL-1998 (first entry)
XX DE Picornavirus Ljungan 145SL partial structural protein.
XX KW Vaccine; prophylactic treatment; myocarditis; multiple sclerosis
XX KW diabetes mellitus; DM; sudden infant death syndrome.
XX OS Picornaviridae.
XX PN WO9811133-A1.
XX PD 19-MAR-1998.
XX PF 09-SEP-1997; 97WO-SE001515.
XX PR 11-SEP-1996; 96SE-00003305.
XX (NIK1/) NIKLASSON B.
XX PI Niklasson B;
XX DR WPI; 1998-207327/18.
XX New picornavirus group causing mammalian disease - useful for d
XX vaccine and medicine development, for treating or preventing e;
XX myocarditis, multiple sclerosis, diabetes mellitus etc.
XX Claim 4; Page 29; 37pp; English.
XX The present invention provides for a new group of picornaviruses;
XX picornaviruses were isolated from bank voles. The present parti
XX structural protein is encoded by the polprotein encoding cDNA of
XX Picornavirus Ljungan 145SL. The invention also claims that vacci
XX be prepared which include, as an immunising or neutralising comp
XX the picornavirus (optionally in attenuated or "killed" form), ar
XX including a subunit of the virus or DNA corresponding to the vir
XX genomic RNA. Medicaments including one of these components as ar
XX ingredient are claimed to be useful for prophylactic or therapeu
XX treatment of diseases caused by the picornavirus in mammals, esp
XX humans, e.g. myocarditis, multiple sclerosis, diabetes mellitus
XX sudden infant death syndrome. (Updated on 27-AUG-2003 to correct

CC AAY36224 to AAY36727 represent the secreted proteins encoded b
 CC human genes. The genes and their corresponding secreted polype
 CC useful for preventing, treating or ameliorating medical condit
 CC by protein or gene therapy. Also pathological conditions can b
 CC determining the amount of the new polypeptides in a sample
 CC determining the presence of mutations in the new genes. Specif
 CC described for each of the 110 genes, based on which tissues th
 CC highly expressed in, and include developing products for the d
 CC treatment of cancer, tumours, developmental abnormalities and
 CC deficiencies, blood disorders, diseases of the immune system,
 CC diseases, inflammation, allergies, Alzheimer's and cognitive d
 CC schizophrenia, arthritis, asthma, psoriasis, sepsis, skin diso
 CC atherosclerosis, diabetes, cardiovascular disorders, kidney di
 CC digestive/endocrine disorders, infections and AIDS. The polype
 CC also useful for identifying their binding partners. The sequen
 CC in AAX97907 to AAX97915 and AAY36223 are used in the exemplifi
 CC the present invention
 CC
 XX SQ Sequence 23 AA;

Query Match 3.9%; Score 7; DB 2; Length 23;
 Best Local Similarity 100.0%; Pred. No. 7.7;
 Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 41 DGGIHL 47
 |||||
 Db 4 DGGIHL 10

RESULT 3
 ADAL1922
 ID ADAL1922 standard; protein; 23 AA.
 XX AC ADAL1922;
 XX DT 06-NOV-2003 (first entry)
 XX Human novel secreted protein associated polypeptide #192.
 DE cancer; inflammation; immune disorder; neurological disorder;
 KW blood clotting disorder; food additive; food preservative;
 KW storage capability; fat content; nutritional component; human;
 KW secreted protein.
 XX OS Homo sapiens.
 XX US2003055236-A1.
 XX PD 20-MAR-2003.
 XX PF 14-MAR-2002; 2002US-00097065.
 XX PR 18-DEC-1997; 97US-0068006P.
 XX PR 18-DEC-1997; 97US-0068007P.
 XX PR 18-DEC-1997; 97US-0068008P.
 XX PR 18-DEC-1997; 97US-0068053P.
 XX PR 18-DEC-1997; 97US-0068054P.
 XX PR 18-DEC-1997; 97US-0068057P.
 XX PR 18-DEC-1997; 97US-0068064P.
 XX PR 18-DEC-1997; 97US-0070923P.
 XX PR 19-DEC-1997; 97US-0068169P.
 XX PR 19-DEC-1997; 97US-0068365P.
 XX PR 19-DEC-1997; 97US-0068367P.
 XX PR 19-DEC-1997; 97US-0068368P.
 XX PR 19-DEC-1997; 97US-0068369P.
 XX PR 17-DEC-1998; 98WO-US027059.
 XX PR 17-JUN-1999; 99US-00334595.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Moore PA, Ruben SM, Carter KC, Shi Y, Rosen CA, Soppet DR;
 PI Kyaw H, Wei Y, Florence KA, Duan DR, Florence C, Greene JM
 PI Ferrie AM, Yu G, Janat F, Ni J;

79 AA;
 milarity 100.0%; Score 179; DB 2; Length 179;
 Conservativeness 0; Mismatches 0; Indels 0; Gaps 0;
 OLMEIARMPVYKGERTEPGTNGYFQWSTHSPINWVFDGGIHLDPNLFSSCYN 60
 OLMEIARMPVYKGERTEPGTNGYFQWSTHSPINWVFDGGIHLDPNLFSSCYN 60
 ARGSTVLKLTYYASTENKGLRMAFFPIMMGGTQKQKHKLFLVYCDIGLNNTEMTIPY 120
 ARGSTVLKLTYYASTENKGLRMAFFPIMMGGTQKQKHKLFLVYCDIGLNNTEMTIPY 120
 AGNWMRPTGSGVIGLRLIDVNLRTYNSSSPNAVNCILOVMGNDAKFVPTTSNIW 179
 AGNWMRPTGSGVIGLRLIDVNLRTYNSSSPNAVNCILOVMGNDAKFVPTTSNIW 179

standard; protein; 23 AA.

) (first entry)

human secreted protein encoded by gene 53.

secreted protein; cancer; tumour; developmental abnormality;
 deficiency; blood disorder; immune system disorder; inflammation;
 disease; allergy; Alzheimer's disease; cognitive disorder;
 dia; arthritis; asthma; psoriasis; sepsis; skin disorder;
 osis; diabetes; cardiovascular disorder; kidney disorder;
 disorder; endocrine disorder; infection; AIDS.

is.

is.

98WO-US027059.
 97US-0068006P.
 97US-0068007P.
 97US-0068008P.
 97US-0068053P.
 97US-0068054P.
 97US-0068057P.
 97US-0068064P.
 97US-0070923P.
 97US-0068169P.
 97US-0068365P.
 97US-0068367P.
 97US-0068368P.
 97US-0068369P.

AN GENOME SCI INC.

Ruben SM, Carter KC, Shi Y, Rosen CA, Soppet DR;
 i Y, Florence K, Duan RD, Florence C, Greene JM, Feng P;
 Yu G, Janat F, Ni J;
 18749/35.

d human genes encoding secreted polypeptides.

Page 516; 537pp; English.

AAX98029 represent 110 isolated human secreted protein genes.

57105/53.

HKABT24 nucleic acid molecules and polypeptides, useful for treating, or ameliorating a medical condition, such as inflammation, immune disorders, neurological and blood clotting

Page 26; 118pp; English.

on relates to an isolated HKABT24 nucleic acid molecule. The a, nucleic acids and antibodies are useful for diagnosing a condition or a susceptibility to a pathological condition, ing, treating, or ameliorating a medical condition, such as inflammation and other immune disorders, neurological and blood disorders. The nucleic acids are also useful for chromosome ion, radiation hybrid mapping or long-range restriction a polypeptides and antibodies are useful for providing al probes for differential identification of the tissues chemistry assays. The polypeptide, polynucleotide, agonist or may also be used as a food additive or preservative to decrease storage capabilities, fat content or other components. The present sequence represents the amino acid a novel human secreted protein associated polypeptide. Note: a data for this patent did not form part of the printed on but was obtained in electronic format directly from USPTO aspto.gov.uk/sequence.html?DocID=20030055236.

AA;

3.9%; Score 7; DB 6; Length 23;

ilarity 100.0%; Pred. No. 7.7;

Conservative 0; Mismatches 0; Indels 0; Gaps 0;

3HLE 47

|||||

3HLE 10

undard; protein; 26 AA.

(first entry)

human secreted protein encoded by gene 53.

ited protein; cancer; tumour; developmental abnormality; :ency; blood disorder; immune system disorder; inflammation; disease; allergy; Alzheimer's disease; cognitive disorder; a; arthritis; asthma; psoriasis; sepsis; skin disorder; psis; diabetes; cardiovascular disorder; kidney disorder; .sorder; endocrine disorder; infection; AIDS.

98WO-US027059.

97US-0068006P.

97US-0068007P.

97US-0068008P.

97US-0068009P.

97US-0068054P.

97US-0068057P.

97US-0068064P.

97US-0070923P.

97US-0068169P.

PR 19-DEC-1997; 97US-0068365P.
PR 19-DEC-1997; 97US-0068367P.
PR 19-DEC-1997; 97US-0068368P.
PR 19-DEC-1997; 97US-0068369P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Moore PA, Ruben SM, Carter KC, Shi Y, Rosen CA, Soppet DR;
PI Kyaw H, Wei Y, Florence K, Duan RD, Florence C, Greene JM,
PI Ferrie AM, Yu G, Janat F, Ni J;
XX
XX WPI; 1999-418749/35.
XX
XX New isolated human genes encoding secreted polypeptides.
XX
XX Disclosure; Page 510; 537pp; English.

CC AAX97916 to AAX98029 represent 110 isolated human secreted prot
CC AAX36224 to AAX36727 represent the secreted proteins encoded by
CC human genes. The genes and their corresponding secreted polypep
CC useful for preventing, treating or ameliorating medical conditi
CC by protein or gene therapy. Also pathological conditions can be
CC by determining the amount of the new polypeptides in a sample o
CC determining the presence of mutations in the new genes. Specific
CC described for each of the 110 genes, based on which tissues the
CC highly expressed in, and include developing products for the di
CC treatment of cancer, tumours, developmental abnormalities and f
CC deficiencies, blood disorders, diseases of the immune system, a
CC diseases, inflammation, allergies, Alzheimer's and cognitive di
CC schizophrenia, arthritis, asthma, psoriasis, sepsis, skin disor
CC atherosclerosis, diabetes, cardiovascular disorders, kidney dis
CC digestive/endocrine disorders, infections and AIDS. The polypep
CC also useful for identifying their binding partners. The sequenc
CC in AAX97907 to AAX97915 and AAX36223 are used in the exemplific
CC the present invention

SQ Sequence 26 AA;

Query Match 3.9%; Score 7; DB 2; Length 26;

Best Local Similarity 100.0%; Pred. No. 8.5;

Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 41 DGGIHLE 47

|||||

Db 8 DGGIHLE 14

RESULT 5

ADAL1903

ID ADAL1903 standard; protein; 26 AA.

XX

AC ADAL1903;

XX

DT 06-NOV-2003 (first entry)

XX

DE Human novel secreted protein associated polypeptide #173.

XX

KW cancer; inflammation; immune disorder; neurological disorder;

KW blood clotting disorder; food additive; food preservative;

KW storage capability; fat content; nutritional component; human;

KW secreted protein.

XX

OS Homo sapiens.

XX

PN US2003055236-A1.

XX

PD 20-MAR-2003.

XX

PF 14-MAR-2002; 2002US-00097065.

XX

PR 18-DEC-1997; 97US-0068006P.

PR 18-DEC-1997; 97US-0068007P.

PR 18-DEC-1997; 97US-0068008P.

97US-0068053P.
97US-0068054P.
97US-0068057P.
97US-0068064P.
97US-0070923P.
97US-0068169P.
97US-0068365P.
97US-0068367P.
97US-0068368P.
97US-0068369P.
98WO-US027059.
99US-00334595.

AN GENOME SCI INC.

Ruben SM, Carter KC, Shi Y, Rosen CA, Soppet DR;
i Y, Florence KA, Duan DR, Florence C, Greene JM, Feng P;
Yu G, Janat F, Ni J;

67105/53.

3 HKABT24 nucleic acid molecules and polypeptides, useful for
treating, or ameliorating a medical condition, such as
lamination, immune disorders, neurological and blood clotting

Page 25; 118pp; English.

on relates to an isolated HKABT24 nucleic acid molecule. The
s, nucleic acids and antibodies are useful for diagnosing a
i condition or a susceptibility to a pathological condition,
ing, treating, or ameliorating a medical condition, such as
lamination and other immune disorders, neurological and blood
sorders. The nucleic acids are also useful for chromosome
ion, radiation hybrid mapping or long-range restriction
e polypeptides and antibodies are useful for providing
al probes for differential identification of the tissues
chemistry assays. The polypeptide, polynucleotide, agonist or
may also be used as a food additive or preservative to
decrease storage capabilities, fat content or other
components. The present sequence represents the amino acid
a novel human secreted protein associated polypeptide. Note:
a data for this patent did not form part of the printed
on but was obtained in electronic format directly from USPTO
uspto.gov.uk/sequence.html?DocID=20030055236.

AA;

ilarity 3.9%; Score 7; DB 6; Length 26;
Conservative 0; Mismatches 0; Indels 0; Gaps 0;

SIHLE 47
|||||
SIHLE 14

andard; protein; 72 AA.

(first entry)

783 encoded by probe for measuring placental gene expression.
oarray; human; placenta; antenatal diagnosis;
order.

s.

-A2.

XX 09-AUG-2001.
PD 30-JAN-2001; 2001WO-US000663.
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
XX Human genome-derived single exon nucleic acid probes useful for
PT gene expression in human placenta.
XX Claim 27; SEQ ID NO 38015; 654pp; English.
XX The present invention relates to single exon nucleic acid probe
CC see AAI31315-AAI57546). The present sequence is a peptide encod
CC such probe. The probes are useful for producing a microarray fo
CC predicting, measuring and displaying gene expression in samples
CC from human placenta. The probes are useful for antenatal diagno
CC human genetic disorders
XX Sequence 72 AA;
Query Match 3.9%; Score 7; DB 4; Length 72;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0;
QY 14 KGERTEP 20
DB 28 KGERTEP 34
RESULT 7
AAAM77564
ID AAAM77564 standard; protein; 72 AA.
XX AAAM77564;
AC AAAM77564;
XX 06-NOV-2001 (first entry)
XX Human bone marrow expressed probe encoded protein SEQ ID NO: 37
DE Human bone marrow expressed exon; gene expression analysis; pr
KW Human; bone marrow expressed exon; gene expression analysis; pr
KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX Homo sapiens.
XX WO200157276-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000668.
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX

unzel DK, Chen W, Rank DR;
18900/53.
r-derived single exon nucleic acid probes useful for analyzing
tion in human bone marrow.
SEQ ID NO 37870; 658pp + Sequence Listing; English.
invention provides a number of single exon nucleic acid
are derived from genomic sequences expressed in the human
They can be used to measure gene expression in bone marrow
ch may enable the improved diagnosis and treatment of cancers
homa, leukaemia and myeloma. The present sequence is a
ded by one of the probes of the invention
AA;
3.9%; Score 7; DB 4; Length 72;
larity 100.0%; Pred. No. 21;
Conservative 0; Mismatches 0; Indels 0; Gaps 0;
RTEP 20
|||||
RTEP 34
undard; protein; 72 AA.
(first entry)
expressed single exon probe encoded protein SEQ ID NO: 36913.
expressed exon; gene expression analysis; probe; microarray;
disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
i.
A2.
2001WO-US000667.
2000US-0180312P.
2000US-0207456P.
2000US-00608408.
2000US-00632366.
2000US-0234687P.
2000US-0236359P.
2000GB-00024263.
XUCULAR DYNAMICS INC.
unzel DK, Chen W, Rank DR;
13446/52.
nucleic acid probes for analyzing gene expression in human
SEQ ID NO 36913; 650pp + Sequence Listing; English.
invention provides a number of single exon nucleic acid
are derived from genomic sequences expressed in the human
can be used to measure gene expression in brain cell samples,
able the diagnosis and improved treatment of nervous system
ch as Alzheimer's disease, multiple sclerosis, schizophrenia,
i cancers. The present sequence is a protein encoded by one of

CC the probes of the invention
XX
SQ Sequence 72 AA;
Query Match 3.9%; Score 7; DB 4; Length 72;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0;
QY 14 KGERTEP 20
Db 28 KGERTEP 34
RESULT 9
ABG59207
ID ABG59207 standard; peptide; 72 AA.
XX
AC ABG59207;
XX
DT 25-FEB-2003 (first entry)
XX
DE Human liver peptide, SEQ ID NO 37855.
XX
XX
KW Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia
KW hypercholesterolaemia; coronary heart disease.
XX
OS Homo sapiens.
XX
PN WO200157273-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000664.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488898/53.
DR Human genome-derived single exon nucleic acid probes useful for
XX gene expression in human adult liver.
XX
PS Claim 27; SEQ ID NO 37855; 658pp; English.
XX
CC The invention relates to a single exon nucleic acid probe (SENP
CC measuring human gene expression in a sample derived from human
CC liver, comprising one of 13109 defined nucleotide sequences giv
CC specification (or complements/ fragments). The probe hybridises
CC stringency to a nucleic acid molecule expressed in the human ad
CC (1) may be used for predicting, measuring and displaying gene e
CC in samples derived from human adult liver. The genes identified
CC involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia;
CC associated with coronary heart disease. ABG47348-ABG59930 repre
CC liver single exon encoded peptides of the invention. Note: The
CC information for this patent does not appear in the printed spec
CC but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 72 AA;
Query Match 3.9%; Score 7; DB 4; Length 72;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0;

ERTEP 20
|||||
ERTEP 34

andard; peptide; 72 AA.

(first entry)

de encoded by genome-derived single exon probe SEQ ID 36257.

le exon probe; asthma; lung cancer; COPD; ILD;
ructive pulmonary disease; interstitial lung disease;
opathic pulmonary fibrosis; neurofibromatosis;
lerosis; Gaucher's disease; Niemann-Pick disease;
ulak syndrome; sarcoidosis; pulmonary haemosiderosis;
istiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
lveolar proteinosis; fibrocystic pulmonary dysplasia;
lary dyskinesia; pulmonary hypertension;
brane disease.

S.

-A2.

; 2001WO-US0000665.

; 2000US-0180312P.
; 2000US-0207456P.
; 2000US-00608408.
; 2000US-00632366.
; 2000US-0234687P.
; 2000US-0236359P.
; 2000GB-00024263.

ECULAR DYNAMICS INC.

anzel DK, Chen W, Rank DR;

14183/15.

dressable set of single exon nucleic acid probes, used to
e expression in human lung samples.

EQ ID NO 36257; 634pp; English.

on relates to a spatially-addressable set of single exon
d probes for measuring gene expression in a sample derived
lung comprising single exon nucleic acid probes having one of
ic acid sequences mentioned in the specification, or their
or the 12387 open reading frames derived from the 12614
o included are a microarray comprising the novel set of probes
set of probes which hybridise at high stringency to a nucleic
sed in the human lung; measuring gene expression in a sample
m human lung, comprising (a) contacting the array with a
of detectably labeled nucleic acids derived from human lung
b) measuring the label detectably bound to each probe of the
tifying exons in a eukaryotic genome, comprising (a)
ally predicting at least one exon from genomic sequences of
te; and (b) detecting specific hybridisation of detectably
leic acids from eukaryotic lung mRNA, to a single exon probe,
agment identical to the predicted exon, the probe is included
e mentioned microarray; assigning exons to a single gene.
a) identifying exons from genomic sequence by the method
b) measuring the expression of each of the exons in several
/or cell types using hybridisation to a single exon

CC microarrays having a probe with the exon, where a common pattern
CC expression of the exons in the tissues and/or cell types indica
CC the exons should be assigned to a single gene; a peptide compri
CC of 12011 sequences, mentioned in the specification, or encoded
CC probes/open reading frames (ORF). The probes are used for gene
CC analysis, and for identifying exons in a gene, particularly usi
CC lung derived mRNA and for the study of lung diseases such as
CC cancer, chronic obstructive pulmonary disease (COPD), interstit
CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofit
CC tuberos scleriosis, Gaucher's disease, Niemann-Pick disease, He
CC Pudiak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmo
CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar prot
CC Karagener syndrome, fibrocystic pulmonary dysplasia, primary ci
CC dyskinesia, pulmonary hypertension and hyaline membrane disease
CC present sequence is a peptide/protein encoded by a single exon
CC the invention. Note: The sequence data for this patent did not
CC of the printed specification, but was obtained in electronic fo
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 72 AA;

Query Match 3.9%; Score 7; DB 5; Length 72;

Best Local Similarity 100.0%; Pred. No. 21;

Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 14 KGERTEP 20

Db 28 KGERTEP 34

RESULT 11

AAAM15491

ID AAAM15491 standard; protein; 88 AA.

AC AAAM15491;

XX 12-OCT-2001 (first entry)

DE Peptide #1925 encoded by probe for measuring cervical gene expr
XX Probe; human; microarray; gene expression; cervical epithelial
KW cervical cancer.

XX Homo sapiens.

XX WO200157278-A2.

XX 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US000670.

XX 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488901/53.

XX Human genome-derived single exon nucleic acid probes useful for
PT gene expression in human cervical epithelial cells.

PS Claim 27; SEQ ID NO 20317; 487pp; English.

CC The present invention relates to human single exon nucleic acid
CC (SENP: see AA10068-AA128459). The present sequence is a peptid
CC by one such probe. The SENPs are derived from human HeLa cells.

to produce a single exon microarray, which can be used for human gene expression in a sample derived from human cervical cells. By measuring gene expression, the probes are therefore useful for staging of diseases of the cervix, notably cancer. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format from WIPO at ftp.wipo.int/pub/published_pct_sequences

AA;

3.9%; Score 7; DB 4; Length 88;
 Similarity 100.0%; Pred. No. 26;
 Conservative 0; Mismatches 0; Indels 0; Gaps 0;

HHLE 47
 |||||
 HHLE 15

undard; protein; 88 AA.

(first entry)

.6 encoded by probe for measuring placental gene expression.

array; human; placenta; antenatal diagnosis;

i.

A2.

2001WO-US000663.

2000US-0180312P.
 2000US-0207456P.
 2000US-00608408.
 2000US-00632366.
 2000US-0234687P.
 2000US-0236359P.
 2000GB-00024263.

MOLECULAR DYNAMICS INC.

anzel DK, Chen W, Rank DR;

18897/53.

1-derived single exon nucleic acid probes useful for analyzing human placenta.

SEQ ID NO 28248; 654pp; English.

invention relates to single exon nucleic acid probes (SENPs: AA157546). The present sequence is a peptide encoded by one of the probes useful for producing a microarray for measuring and displaying gene expression in samples derived from placenta. The probes are useful for antenatal diagnosis of genetic disorders

AA;

3.9%; Score 7; DB 4; Length 88;
 Similarity 100.0%; Pred. No. 26;
 Conservative 0; Mismatches 0; Indels 0; Gaps 0;

HHLE 47

|||||
 9 DGGIHLE 15

Db

RESULT 13

ABB29325
 ID ABB29325 standard; peptide; 88 AA.

XX

AC ABB29325;

XX 01-FEB-2002 (first entry)

XX Peptide #1976 encoded by breast cell single exon nucleic acid probe;
 XX Human; microarray; single exon probe; gene expression; breast;
 XX cancer.

XX Homo sapiens.

XX WO200157271-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000662.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-496933/54.

XX New spatially-addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived from human breast and BT 474 cells. The method involves contacting probes with a collection of detectably labelled nucleic acids derived from mRNA of human breast, and then measuring the label bound to the probe of the microarray. The probes are useful for verifying the expression of regions of genomic DNA predicted to encode proteins are useful for gene discovery, and for determining predisposition to disease. Gene expression analysis is useful for assessing the toxicity of chemical agents on cells. The microarray invention presents a far greater diversity of probes for measuring gene expression, with far less bias than expressed sequence tag microarrays. The method is suitable for rapid production of functional information from genomic sequence. The present sequence is a peptide encoded by a single exon nucleic acid probe of the invention. No sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from at ftp.wipo.int/pub/published_pct_sequences

Claim 27; SEQ ID NO 12293; 327pp + Sequence Listing; English.

XX The invention relates to a spatially-addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived from human breast and BT 474 cells. The method involves contacting probes with a collection of detectably labelled nucleic acids derived from mRNA of human breast, and then measuring the label bound to the probe of the microarray. The probes are useful for verifying the expression of regions of genomic DNA predicted to encode proteins are useful for gene discovery, and for determining predisposition to disease. Gene expression analysis is useful for assessing the toxicity of chemical agents on cells. The microarray invention presents a far greater diversity of probes for measuring gene expression, with far less bias than expressed sequence tag microarrays. The method is suitable for rapid production of functional information from genomic sequence. The present sequence is a peptide encoded by a single exon nucleic acid probe of the invention. No sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 88 AA;

Query Match 3.9%; Score 7; DB 4; Length 88;

Best Local Similarity 100.0%; Pred. No. 26;

Matches 7; Conservative 0; Mismatches 0; Indels 0;

Qy 41 DGGIHLE 47

Db 9 DGGIHLE 15

06-NOV-2001 (first entry)
 Human bone marrow expressed probe encoded protein SEQ ID NO: 27
 Human; bone marrow expressed exon; gene expression analysis; pr
 microarray; cancer; leukaemia; lymphoma; myeloma.
 Homo sapiens.
 WO200157276-A2.
 09-AUG-2001.
 30-JAN-2001; 2001WO-US000668.
 04-FEB-2000; 2000US-0180312P.
 26-MAY-2000; 2000US-0207456P.
 30-JUN-2000; 2000US-00608408.
 03-AUG-2000; 2000US-00632366.
 21-SEP-2000; 2000US-0234687P.
 27-SEP-2000; 2000US-0236359P.
 04-OCT-2000; 2000GB-00024263.
 (MOLE-) MOLECULAR DYNAMICS INC.
 Penn SG, Hanzel DK, Chen W, Rank DR;
 WPI; 2001-488900/53.
 Human genome-derived single exon nucleic acid probes useful for
 gene expression in human bone marrow.
 Example 4; SEQ ID NO 27988; 658pp + Sequence Listing; English.
 The present invention provides a number of single exon nucleic
 probes which are derived from genomic sequences expressed in th
 bone marrow. They can be used to measure gene expression in bon
 samples, which may enable the improved diagnosis and treatment
 such as lymphoma, leukaemia and myeloma. The present sequence i
 protein encoded by one of the probes of the invention

07 encoded by probe for measuring heart cell gene expression.
 expression; heart; microarray; vascular system;
 lar disease; hypertension; cardiac arrhythmia;
 heart disease.
 S.
 -A2.
 2001WO-US000666.
 2000US-0180312P.
 2000US-0207456P.
 2000US-00608408.
 2000US-00632366.
 2000US-0234687P.
 2000US-0236359P.
 2000GB-00024263.
 MOLECULAR DYNAMICS INC.
 unzel DK, Chen W, Rank DR;
 88999/53.
 nucleic acid probes for analyzing gene expression in human
 invention relates to single exon nucleic acid probes for
 man gene expression in a sample derived from human heart (see
 1305). The present sequence is a protein encoded by one such
 probes may be used for predicting, measuring and displaying
 ion in samples derived from the human heart via microarrays.
 gene expression, the probes are useful for predicting,
 grading, staging, monitoring and prognosing diseases of the
 and vascular system e.g. cardiovascular disease,
 cardiac arrhythmias and congenital heart disease. Note: The
 a for this patent did not form part of the printed
 n, but was obtained in electronic format directly from WIPO
 int/pub/published_pct_sequences

Query Match 3.9%; Score 7; DB 4; Length 88;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 7; Conservative 0; Mismatches 0; Indels 0;
 Qy 41 DGGHLE 47
 Db 9 DGGHLE 15
 Search completed: April 23, 2004, 14:06:54
 Job time : 62 secs

3.9%; Score 7; DB 4; Length 88;
 larity 100.0%; Pred. No. 26;
 Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 HLE 47
 HLE 15
 undard; protein; 88 AA.

09:21:14 2004

us-09-147-801d-4.Oligo.ra1

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

main search, using sw model

April 23, 2004, 14:05:45 ; Search time 23 Seconds
(without alignments)
401.785 Million cell updates/sec

IS-09-147-801D-4
KDLMEIARMPSEVYKGERTEP.....QVKMGNDKFMVPTTSNIVW 179

hapop 60.0, Gapext 60.0

19414 seqs, 51625971 residues

0
uts satisfying chosen parameters: 389414

length: 0

length: 2000000000

Listing first 45 summaries

Issued Patents AA.*

:/cgn2_6/ptodata/2/iaa/5A_COMB.pap.*

:/cgn2_6/ptodata/2/iaa/5B_COMB.pap.*

:/cgn2_6/ptodata/2/iaa/6A_COMB.pap.*

:/cgn2_6/ptodata/2/iaa/6B_COMB.pap.*

:/cgn2_6/ptodata/2/iaa/PCTUS_COMB.pap.*

:/cgn2_6/ptodata/2/iaa/backfiles.pap.*

s the number of results predicted by chance to have a
er than or equal to the score of the result being printed,
ved by analysis of the total score distribution.

SUMMARIES

Query	Length	DB	ID	Description
3.9	290	4	US-09-910-174B-19	Sequence 19, Appl
3.9	290	4	US-09-620-461-19	Sequence 19, Appl
3.9	319	4	US-09-910-174B-12	Sequence 12, Appl
3.9	319	4	US-09-620-461-12	Sequence 12, Appl
3.9	342	2	US-08-724-394A-6	Sequence 6, Appli
3.9	350	4	US-09-651-200-25	Sequence 25, Appl
3.9	350	4	US-09-910-174B-17	Sequence 17, Appl
3.9	350	4	US-08-620-461-17	Sequence 17, Appl
3.9	357	4	US-09-910-174B-14	Sequence 14, Appl
3.9	357	4	US-09-620-461-14	Sequence 14, Appl
3.9	474	4	US-08-134-001C-3176	Sequence 3176, Ap
3.9	504	4	US-09-252-991A-19852	Sequence 19852, A
3.9	513	4	US-09-910-174B-18	Sequence 18, Appl
3.9	513	4	US-09-620-461-18	Sequence 18, Appl
3.9	540	2	US-08-724-394A-4	Sequence 4, Appli
3.9	584	4	US-09-910-174B-16	Sequence 16, Appl
3.9	584	4	US-09-620-461-16	Sequence 16, Appl
3.9	610	2	US-08-724-394A-5	Sequence 5, Appli
3.9	621	2	US-08-419-652-4	Sequence 4, Appli
3.9	660	1	US-08-248-532-3	Sequence 3, Appli
3.9	660	2	US-08-419-652-3	Sequence 3, Appli
3.9	662	1	US-08-248-532-2	Sequence 2, Appli
3.9	662	2	US-08-419-652-2	Sequence 2, Appli
3.9	662	2	US-08-685-118-4	Sequence 4, Appli
3.9	662	2	US-08-915-495-4	Sequence 4, Appli
3.9	662	2	US-08-914-520-4	Sequence 4, Appli
3.9	662	3	US-08-789-350-2	Sequence 2, Appli

28	7	3.9	731	4	US-09-910-174B-15	Sequence
29	7	3.9	731	4	US-09-620-461-15	Sequence
30	6	3.4	38	1	US-08-118-270-192	Sequence
31	6	3.4	38	5	PCT-US93-08528-192	Sequence
32	6	3.4	49	4	US-09-227-357-595	Sequence
33	6	3.4	86	4	US-09-489-039A-8767	Sequence
34	6	3.4	90	4	US-09-134-001C-3797	Sequence
35	6	3.4	110	4	US-09-205-258-811	Sequence
36	6	3.4	130	4	US-09-252-991A-23635	Sequence
37	6	3.4	137	3	US-09-431-480-18	Sequence
38	6	3.4	137	3	US-09-617-302-18	Sequence
39	6	3.4	148	4	US-09-489-039A-9136	Sequence
40	6	3.4	172	4	US-09-621-976-4131	Sequence
41	6	3.4	174	4	US-09-149-476-635	Sequence
42	6	3.4	176	4	US-09-328-352-7007	Sequence
43	6	3.4	233	4	US-09-489-039A-9119	Sequence
44	6	3.4	236	4	US-09-252-991A-21645	Sequence
45	6	3.4	268	1	US-08-446-919A-2	Sequence

ALIGNMENTS

RESULT 1
US-09-910-174B-19
; Sequence 19, Application US/09910174B
; Patent No. 6630575
; GENERAL INFORMATION:
; APPLICANT: Coyle, Anthony J.
; APPLICANT: Fraser, Christopher C.
; APPLICANT: Manning, Stephen
; TITLE OF INVENTION: B7-H2 Molecules, No. 6630575el Members of the
; TITLE OF INVENTION: Family and Uses Thereof
; FILE REFERENCE: 35800/236924
; CURRENT APPLICATION NUMBER: US/09/910,174B
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US 09/620,461
; PRIOR FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-910-174B-19

Query Match 3.9%; Score 7; DB 4; Length 290;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0;

Qy 41 DGGTHLE 47
Db 157 DGGHLE 163

RESULT 2
US-09-620-461-19
; Sequence 19, Application US/09620461
; Patent No. 6635750
; GENERAL INFORMATION:
; APPLICANT: Coyle, Anthony J.
; APPLICANT: Fraser, Christopher C.
; APPLICANT: Manning, Stephen
; TITLE OF INVENTION: B7-H2 Molecules, No. 6635750el Members of the
; TITLE OF INVENTION: Family and Uses Thereof
; FILE REFERENCE: 5800-149
; CURRENT APPLICATION NUMBER: US/09/620,461
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 19
; LENGTH: 290
; TYPE: PRT

09:21:14 2004

us-09-147-801d-4.Oligo.ra1

no sapiens

3.9%; Score 7; DB 4; Length 290;
ilarity 100.0%; Pred. No. 19;
Conservative 0; Mismatches 0; Indels 0; Gaps 0;

3HLE 47
|||||
3HLE 163

Application US/09910174B
575

le, Anthony J.
aser, Christopher C.

inning, Stephen
TION: B7-H2 Molecules, No. 6630575el Members of the B7
TION: Family and Uses Thereof

3: 35800/236924

ATION NUMBER: US/09/910,174B

DATE: 2001-07-20

ION NUMBER: US 09/620,461

DATE: 2000-07-20

ID NOS: 32

SEQ for Windows Version 4.0

no sapiens

3.9%; Score 7; DB 4; Length 319;
ilarity 100.0%; Pred. No. 21;
Conservative 0; Mismatches 0; Indels 0; Gaps 0;

3HLE 47
|||||
3HLE 165

plication US/09620461

750

TION:

yle, Anthony J.

aser, Christopher C.

inning, Stephen

TION: B7-H2 Molecules, No. 6635750el Members of the B7

TION: Family and Uses thereof

3: 5800-149

ATION NUMBER: US/09/620,461

DATE: 2000-07-20

ID NOS: 29

tSEQ for Windows Version 3.0

no sapiens

3.9%; Score 7; DB 4; Length 319;
ilarity 100.0%; Pred. No. 21;
Conservative 0; Mismatches 0; Indels 0; Gaps 0;

3HLE 47
|||||
3HLE 165

RESULT 5

US-08-724-394A-6

; Sequence 6, Application US/08724394A

; Patent No. 5872237

; GENERAL INFORMATION:

; APPLICANT: Feder, John N.

; APPLICANT: Krommal, Gregory S.

; APPLICANT: Lauer, Peter M.

; APPLICANT: Ruddy, David A.

; APPLICANT: Thomas, Winston

; APPLICANT: Tsuchihashi, Zenta

; APPLICANT: Wolff, Roger K.

; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el

; TITLE OF INVENTION: Sequences and Antibodies Thereto

; NUMBER OF SEQUENCES: 31

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP

; STREET: Two Embarcadero Center, 8th Floor

; CITY: San Francisco

; STATE: CA

; COUNTRY: USA

; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/724,394A

; FILING DATE: 01-OCT-1996

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: Fitts, Renee A.

; REGISTRATION NUMBER: 35,136

; REFERENCE/DOCKET NUMBER: 017957-000100

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-576-0200

; TELEFAX: 415-576-0300

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 342 amino acids

; TYPE: amino acid

; STRANDEDNESS: not relevant

; TOPOLOGY: not relevant

; MOLECULE TYPE: peptide

; FEATURE:

; NAME/KEY: Region

; LOCATION: 1..342

; OTHER INFORMATION: /note= "BTF4"

; US-08-724-394A-6

Query Match 3.9%; Score 7; DB 2; Length 342;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 41 DGGIHLE 47

Db 162 DGGIHLE 168

RESULT 6

US-09-651-200-25

; Sequence 25, Application US/09651200

; Patent No. 6429303

; GENERAL INFORMATION:

; APPLICANT: Green et al

; TITLE OF INVENTION: Polynucleotides Encoding Members of the Human

; TITLE OF INVENTION: Lymphocyte Activation Antigen B-7 Family and

; TITLE OF INVENTION: Polypeptides Encoded Thereby

; FILE REFERENCE: 15966-562 (CURA-62)

; CURRENT APPLICATION NUMBER: US/09/651,200

; CURRENT FILING DATE: 2000-08-30

ION NUMBER: 60/152383
 ATE: 1999-09-03
 ION NUMBER: 60/172909
 ATE: 1999-12-21
 ION NUMBER: 60/183578
 ATE: 2000-02-18
 ID NOS: 25
 entIn Ver. 2.0

o sapiens

3.9%; Score 7; DB 4; Length 350;
 larity 100.0%; Pred. No. 23;
 Conservative 0; Mismatches 0; Indels 0; Gaps 0;

IHLE 47
 |||||
 IHLE 163

plication US/09910174B

575
 ION:
 le, Anthony J.
 aser, Christopher C.
 nning, Stephen
 ION: B7-H2 Molecules, No. 6630575el Members of the B7
 ION: Family and Uses Thereof
 : 35800/236924
 ACTION NUMBER: US/09/910,174B
 DATE: 2001-07-20
 ION NUMBER: US 09/620,461
 ATE: 2000-07-20
 ID NOS: 32
 SEQ for Windows Version 4.0

o sapiens

3.9%; Score 7; DB 4; Length 350;
 larity 100.0%; Pred. No. 23;
 Conservative 0; Mismatches 0; Indels 0; Gaps 0;

IHLE 47
 |||||
 IHLE 163

plication US/09620461

750
 ION:
 le, Anthony J.
 aser, Christopher C.
 nning, Stephen
 ION: B7-H2 Molecules, No. 6635750el Members of the B7
 ION: Family and Uses Thereof
 : 5800-149
 ACTION NUMBER: US/09/620,461
 DATE: 2000-07-20
 ID NOS: 29
 SEQ for Windows Version 3.0

; ORGANISM: Homo sapiens
 US-09-620-461-17
 Query Match 3.9%; Score 7; DB 4; Length 350;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 41 DGGIHLE 47
 |||||
 Db 157 DGGIHLE 163

RESULT 9

US-09-910-174B-14
 ; Sequence 14, Application US/09910174B
 ; Patent No. 6630575
 ; GENERAL INFORMATION:
 ; APPLICANT: Coyle, Anthony J.
 ; APPLICANT: Fraser, Christopher C.
 ; APPLICANT: Manning, Stephen
 ; TITLE OF INVENTION: B7-H2 Molecules, No. 6630575el Members of the
 ; TITLE OF INVENTION: Family and Uses Thereof
 ; FILE REFERENCE: 35800/236924
 ; CURRENT APPLICATION NUMBER: US/09/910,174B
 ; CURRENT FILING DATE: 2001-07-20
 ; PRIOR APPLICATION NUMBER: US 09/620,461
 ; PRIOR FILING DATE: 2000-07-20
 ; NUMBER OF SEQ ID NOS: 32
 ; SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO 14
 ; LENGTH: 357
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-910-174B-14

Query Match 3.9%; Score 7; DB 4; Length 357;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 41 DGGIHLE 47
 |||||
 Db 157 DGGIHLE 163

RESULT 10

US-09-620-461-14
 ; Sequence 14, Application US/09620461
 ; Patent No. 6635750
 ; GENERAL INFORMATION:
 ; APPLICANT: Coyle, Anthony J.
 ; APPLICANT: Fraser, Christopher C.
 ; APPLICANT: Manning, Stephen
 ; TITLE OF INVENTION: B7-H2 Molecules, No. 6635750el Members of the
 ; TITLE OF INVENTION: Family and Uses Thereof
 ; FILE REFERENCE: 5800-149
 ; CURRENT APPLICATION NUMBER: US/09/620,461
 ; CURRENT FILING DATE: 2000-07-20
 ; NUMBER OF SEQ ID NOS: 29
 ; SOFTWARE: FastSEQ for Windows Version 3.0
 ; SEQ ID NO 14
 ; LENGTH: 357
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-620-461-14

Query Match 3.9%; Score 7; DB 4; Length 357;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 41 DGGIHLE 47
 |||||
 Db 157 DGGIHLE 163

76
Application US/09134001C
370
TION:
TION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: CTC-007
TION NUMBER: US/09134.001C
: DATE: 1998-08-13
ION NUMBER: US 60/064,964
ATE: 1997-11-08
TION NUMBER: US 60/055,779
ATE: 1997-08-14
ID NOS: 5674

Phylococcus epidermidis

76
3.9%; Score 7; DB 4; Length 474;
larity 100.0%; Pred.No. 30;
Conservative 0; Mismatches 0; Indels 0; Gaps 0;

LNNT 113
||||
LNNT 290

Application US/09252991A

852
795
rc J. Rubenfield et al.
TION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
: 107196.136
TION NUMBER: US/09/252,991A
: DATE: 1999-02-18
ION NUMBER: US 60/074,788
ATE: 1998-02-18
ION NUMBER: US 60/094,190
ATE: 1998-07-27
ID NOS: 33142

Pseudomonas aeruginosa

852
3.9%; Score 7; DB 4; Length 504;
larity 100.0%; Pred.No. 32;
Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RGSV 133
||||
RGSV 292

Application US/09910174B

1575
TION:
TION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
: 107196.136
TION NUMBER: US/09/252,991A
: DATE: 1999-02-18
ION NUMBER: US 60/074,788
ATE: 1998-02-18
ION NUMBER: US 60/094,190
ATE: 1998-07-27
ID NOS: 33142

FILE REFERENCE: 35800/236924
CURRENT APPLICATION NUMBER: US/09/910,174B
CURRENT FILING DATE: 2001-07-20
PRIOR APPLICATION NUMBER: US 09/620,461
PRIOR FILING DATE: 2000-07-20
NUMBER OF SEQ ID NOS: 32
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 18
LENGTH: 513
TYPE: PRT
ORGANISM: Homo sapiens
US-09-910-174B-18

Query Match 3.9%; Score 7; DB 4; Length 513;
Best Local Similarity 100.0%; Pred.No. 32; Indels 0;
Matches 7; Conservative 0; Mismatches 0;

Qy 41 DGGIHL 47
|||||
Db 159 DGGIHL 165

RESULT 14

US-09-620-461-18
Sequence 18, Application US/09620461
Patent No. 6635750
GENERAL INFORMATION:
APPLICANT: Coyle, Anthony J.
APPLICANT: Fraser, Christopher C.
APPLICANT: Manning, Stephen
TITLE OF INVENTION: B7-H2 Molecules, No. 6635750el Members of the
TITLE OF INVENTION: Family and Uses Thereof
FILE REFERENCE: 5800-149
CURRENT APPLICATION NUMBER: US/09/620,461
CURRENT FILING DATE: 2000-07-20
NUMBER OF SEQ ID NOS: 29
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 18
LENGTH: 513
TYPE: PRT
ORGANISM: Homo sapiens
US-09-620-461-18

Query Match 3.9%; Score 7; DB 4; Length 513;
Best Local Similarity 100.0%; Pred.No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0;

Qy 41 DGGIHL 47
|||||
Db 159 DGGIHL 165

RESULT 15

US-08-724-394A-4
Sequence 4, Application US/08724394A
Patent No. 5872237
GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Kronmal, Gregory S.
APPLICANT: Lauer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
TITLE OF INVENTION: Sequences and Antibodies Thereof
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA

09:21:14 2004

us-09-147-801d-4.Oligo.rai

1-3834
TABLE FORM:
E: Floppy disk
IBM PC compatible
SYSTEM: PC-DOS/MS-DOS
Patent In Release #1.0, Version #1.30
ICATION DATA:
N NUMBER: US/08/724,394A
E: 01-OCT-1996
TION: 536
NT INFORMATION:
ts, Renee A.
ON NUMBER: 35,136
DOCKET NUMBER: 017957-000100
ATION INFORMATION:
415-576-0200
415-576-0300
R SEQ ID NO: 4:
RACTERISTICS:
40 amino acids
no acid
SS: not relevant
not relevant
E: peptide
Region
1.540
FORMATION: /note= "BTF5"

3.9%; Score 7; DB 2; Length 540;
larity 100.0%; Pred.No.34;
Conservative 0; Mismatches 0; Indels 0; Gaps 0;

HHLE 47
|||
HHLE 168

April 23, 2004, 14:09:23
S

09:21:14 2004

us-09-147-801d-4.Oligo.rapb

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

in search, using sw model

April 23, 2004, 14:08:20 ; Search time 43 Seconds
(without alignments)
1150.908 Million cell updates/sec

IS-09-147-801D-4
79
KDLMEIARMPVYKGERTEP.....QVKMGNDKAFMPTTSNIV 179

LLIGO

lapop 60.0 , Gapext 60.0

133595 seqs, 276475211 residues

0

its satisfying chosen parameters: 1133595

length: 0

length: 2000000000

Listing first 45 summaries

Published Applications_AA:*
: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

s the number of results predicted by chance to have a
er than or equal to the score of the result being printed,
ved by analysis of the total score distribution.

SUMMARIES

Query	Match	Length	DB	ID	Description
00.0	179	10	US-09-147-801B-4		Sequence 4, Appli
4.5	2179	14	US-10-224-999A-3481		Sequence 3481, Ap
3.9	23	12	US-10-372-876-450		Sequence 450, App
3.9	23	14	US-10-097-065-450		Sequence 450, App
3.9	26	12	US-10-372-876-431		Sequence 431, App
3.9	26	14	US-10-097-065-431		Sequence 431, App
3.9	72	9	US-09-864-761-46085		Sequence 46085, A
3.9	88	9	US-09-864-761-35206		Sequence 35206, A
3.9	93	9	US-09-864-761-36770		Sequence 36770, A
3.9	135	12	US-10-372-876-437		Sequence 437, App
3.9	135	14	US-10-097-065-437		Sequence 437, App
3.9	161	12	US-10-424-599-159096		Sequence 159096,
3.9	220	12	US-10-424-599-183941		Sequence 183941,
3.9	223	14	US-10-197-844-7		Sequence 7, Appli
3.9	245	12	US-10-425-114-44067		Sequence 44067, A

16	7	3.9	252	15	US-10-093-463-166	Sequenc
17	7	3.9	263	9	US-09-955-866-11	Sequence
18	7	3.9	263	9	US-09-896-738-17	Sequence
19	7	3.9	290	9	US-09-910-174A-19	Sequence
20	7	3.9	311	12	US-10-425-114-51302	Sequence
21	7	3.9	319	9	US-09-910-174A-12	Sequence
22	7	3.9	319	14	US-10-156-424A-12	Sequenc
23	7	3.9	324	15	US-10-369-493-8329	Sequenc
24	7	3.9	331	14	US-10-197-844-2	Sequenc
25	7	3.9	334	10	US-09-746-783-134	Sequenc
26	7	3.9	334	12	US-10-463-260-2	Sequenc
27	7	3.9	334	12	US-10-372-876-161	Sequenc
28	7	3.9	334	14	US-10-097-065-161	Sequenc
29	7	3.9	341	15	US-10-093-463-162	Sequenc
30	7	3.9	350	9	US-09-910-174A-17	Sequence
31	7	3.9	356	12	US-10-425-114-53606	Sequence
32	7	3.9	357	9	US-09-910-174A-14	Sequence
33	7	3.9	359	14	US-10-156-424A-11	Sequence
34	7	3.9	390	12	US-10-425-114-52391	Sequence
35	7	3.9	393	12	US-10-087-192-1923	Sequence
36	7	3.9	444	9	US-09-815-242-5482	Sequence
37	7	3.9	473	9	US-09-815-242-12584	Sequence
38	7	3.9	473	12	US-10-282-122A-43964	Sequenc
39	7	3.9	473	12	US-10-282-122A-70714	Sequenc
40	7	3.9	481	12	US-10-463-260-5	Sequenc
41	7	3.9	499	12	US-10-282-122A-71531	Sequenc
42	7	3.9	513	9	US-09-910-174A-18	Sequence
43	7	3.9	584	9	US-09-910-174A-16	Sequence
44	7	3.9	584	9	US-09-955-866-12	Sequence
45	7	3.9	584	9	US-09-896-738-18	Sequence

ALIGNMENTS

RESULT 1
US-09-147-801B-4
; Sequence 4, Application US/09147801B
; Publication No. US20030044960A1
; GENERAL INFORMATION:
; APPLICANT: Niklasson, Bo
; TITLE OF INVENTION: New Picorna Viruses, Vaccines and Diagnostic
; FILE REFERENCE: Niklasson 09/147801
; CURRENT APPLICATION NUMBER: US/09/147,801B
; CURRENT FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: PCT/SE97/01515
; PRIOR FILING DATE: 1997-09-09
; PRIOR APPLICATION NUMBER: SE 9603305-5
; PRIOR FILING DATE: 1996-09-11
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Picornaviridae
US-09-147-801B-4

Query Match 100.0%; Score 179; DB 10; Length 179;
Best Local Similarity 100.0%; Pred. No. 1.8e-177;
Matches 179; Conservative 0; Mismatches 0; Indels 0;
Qy 1 KDLMEIARMPVYKGERTEPGTNGYFQWSHTSPINNVFDDGGTHLEDMNLNI
Db 1 KDLMEIARMPVYKGERTEPGTNGYFQWSHTSPINNVFDDGGTHLEDMNLNI
Qy 61 YWRGSTVLKLTVTYASTFNKGRLMARFPIMMOGTQORKKHKCLFMVCDIGLNTTFE
Db 61 YWRGSTVLKLTVTYASTFNKGRLMARFPIMMOGTQORKKHKCLFMVCDIGLNTTFE
Qy 121 TWGNWMPRTGSGVIGWLRIIDLNLRLTYNSSSPNANVCILQVKMGNDKAFVWPPTTE
Db 121 TWGNWMPRTGSGVIGWLRIIDLNLRLTYNSSSPNANVCILQVKMGNDKAFVWPPTTE

```

i81
Application US/10224999A
US20030171318A1
TION:
iad Genetics, Inc.
rham, Scott
vitz, Kenton
bden, Adrian
TION: Composition and Method for Treating Viral Infection
E: 5004 01
TION NUMBER: US/10/224,999A
DATE: 2003-03-03
TION NUMBER: US 60/313,695
DATE: 2001-08-20
ID NOS: 3484
ntIn version 3.1

ian parechovirus 2
i81

4.5%; Score 8; DB 14; Length 2179;
larity 100.0%; Pred. No. 49;
Conservative 0; Mismatches 0; Indels 0; Gaps 0;

IRLTYN 148
|||||
IRLTYN 510

Application US/10372876
US20030204071A1
TION:
re, Paul A. et al.
TION: 110 Human Secreted Proteins
E: P2021P1
TION NUMBER: US/10/372,876
DATE: 2003-02-26
TION NUMBER: 09/334,595
DATE: 1999-06-17
TION NUMBER: PCT/US98/27059
DATE: 1998-12-17
TION NUMBER: 60/070,923
DATE: 1997-12-18
TION NUMBER: 60/068,007
DATE: 1997-12-18
TION NUMBER: 60/068,057
DATE: 1997-12-18
TION NUMBER: 60/068,006
DATE: 1997-12-18
TION NUMBER: 60/068,369
DATE: 1997-12-19
TION NUMBER: 60/068,367
DATE: 1997-12-19
TION NUMBER: 60/068,368
DATE: 1997-12-19
TION NUMBER: 60/068,169
DATE: 1997-12-19
r Application data removed - See File Wrapper or PALM.
ID NOS: 672
ntIn Ver. 2.0

io sapiens

```

3.9%; Score 7; DB 12; Length 23;

```

Best Local Similarity 100.0%; Pred. No. 10;
Matches 7; Conservative 0; Mismatches 0; Indels 0;

Qy 41 DGGIHLE 47
Db 4 DGGIHLE 10

RESULT 4
US-10-097-065-450
; Sequence 450, Application US/10097065
; Publication No. US20030055236A1
; GENERAL INFORMATION:
; APPLICANT: Moore, Paul A. et al.
; TITLE OF INVENTION: 110 Human Secreted Proteins
; FILE REFERENCE: P2021P1
; CURRENT APPLICATION NUMBER: US/10/097,065
; PRIOR FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: PCT/US98/27059
; PRIOR FILING DATE: 1998-12-17
; PRIOR APPLICATION NUMBER: 60/070,923
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,007
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,057
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,006
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,369
; PRIOR FILING DATE: 1997-12-19
; PRIOR APPLICATION NUMBER: 60/068,367
; PRIOR FILING DATE: 1997-12-19
; PRIOR APPLICATION NUMBER: 60/068,368
; PRIOR FILING DATE: 1997-12-19
; PRIOR APPLICATION NUMBER: 60/068,169
; PRIOR FILING DATE: 1997-12-19
; PRIOR APPLICATION NUMBER: 60/068,053
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,064
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,054
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,008
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,365
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 450
; LENGTH: 23
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-097-065-450

Query Match 3.9%; Score 7; DB 14; Length 23;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 7; Conservative 0; Mismatches 0; Indels 0;

Qy 41 DGGIHLE 47
Db 4 DGGIHLE 10

RESULT 5
US-10-372-876-431
; Sequence 431, Application US/10372876
; Publication No. US20030204071A1
; GENERAL INFORMATION:
; APPLICANT: Moore, Paul A. et al.
; TITLE OF INVENTION: 110 Human Secreted Proteins
; FILE REFERENCE: P2021P1
; CURRENT APPLICATION NUMBER: US/10/372,876
; CURRENT FILING DATE: 2003-02-26

```

ION NUMBER: 09/334,595
 ATE: 1999-06-17
 ION NUMBER: PCT/US98/27059
 ATE: 1998-12-17
 ION NUMBER: 60/070,923
 ATE: 1997-12-18
 ION NUMBER: 60/068,007
 ATE: 1997-12-18
 ION NUMBER: 60/068,057
 ATE: 1997-12-18
 ION NUMBER: 60/068,006
 ATE: 1997-12-18
 ION NUMBER: 60/068,369
 ATE: 1997-12-19
 ION NUMBER: 60/068,367
 ATE: 1997-12-19
 ION NUMBER: 60/068,368
 ATE: 1997-12-19
 ION NUMBER: 60/068,169
 ATE: 1997-12-19
 or Application data removed - See File Wrapper or PALM.
 ID NOS: 672
 ntIn Ver. 2.0

o sapiens

3.9%; Score 7; DB 12; Length 26;
 larity 100.0%; Pred.No.11;
 Conservative 0; Mismatches 0; Indels 0; Gaps 0;

HLHE 47
 ||||
 HLHE 14

pplication US/10097065
 US20030055236A1
 TION:
 re, Paul A. et al.
 : P2021P1
 TION: 110 Human Secreted Proteins
 ATION NUMBER: US/10/097,065
 DATE: 2002-03-14
 ION NUMBER: PCT/US98/27059
 ATE: 1998-12-17
 ION NUMBER: 60/070,923
 ATE: 1997-12-18
 ION NUMBER: 60/068,007
 ATE: 1997-12-18
 ION NUMBER: 60/068,057
 ATE: 1997-12-18
 ION NUMBER: 60/068,006
 ATE: 1997-12-18
 ION NUMBER: 60/068,369
 ATE: 1997-12-19
 ION NUMBER: 60/068,367
 ATE: 1997-12-19
 ION NUMBER: 60/068,368
 ATE: 1997-12-19
 ION NUMBER: 60/068,169
 ATE: 1997-12-19
 ION NUMBER: 60/068,053
 ATE: 1997-12-18
 ION NUMBER: 60/068,064
 ATE: 1997-12-18
 ION NUMBER: 60/068,054
 ATE: 1997-12-18
 ION NUMBER: 60/068,008

; PRIOR FILING DATE: 1997-12-18
 ; PRIOR APPLICATION NUMBER: 60/068,365
 ; PRIOR FILING DATE: 1997-12-19
 ; NUMBER OF SEQ ID NOS: 672
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 431
 ; LENGTH: 26
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-097-065-431

Query Match 3.9%; Score 7; DB 14; Length 26;
 Best Local Similarity 100.0%; Pred.No.11;
 Matches 7; Conservative 0; Mismatches 0; Indels 0;

Qy 41 DGGIHLE 47
 |||||
 Db 8 DGGIHLE 14

RESULT 7

US-09-864-761-46085
 ; Sequence 46085, Application US/09864761
 ; Patent NO. US20020048763A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharron G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; APPLICANT: Chen, Wenheng
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID
 ; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
 ; FILE REFERENCE: Aesomica-X-1
 ; CURRENT APPLICATION NUMBER: US/09/864,761
 ; CURRENT FILING DATE: 2001-05-23
 ; PRIOR APPLICATION NUMBER: US 60/180,312
 ; PRIOR FILING DATE: 2000-02-04
 ; PRIOR APPLICATION NUMBER: US 60/207,456
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: US 09/632,366
 ; PRIOR FILING DATE: 2000-08-03
 ; PRIOR APPLICATION NUMBER: GB 24263.6
 ; PRIOR FILING DATE: 2000-10-04
 ; PRIOR APPLICATION NUMBER: US 60/236,359
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00670
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: US 60/234,687
 ; PRIOR FILING DATE: 2000-09-21
 ; PRIOR APPLICATION NUMBER: US 09/608,408
 ; PRIOR FILING DATE: 2000-06-30
 ; PRIOR APPLICATION NUMBER: US 09/774,203
 ; NUMBER OF SEQ ID NOS: 49117
 ; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 46085

```

no sapiens
ATTN: MAP TO ACO20904.3
ATTN: EXPRESSED IN LUNG, SIGNAL = 3.4
ATTN: EXPRESSED IN PLACENTA, SIGNAL = 1.6
ATTN: EXPRESSED IN BRAIN, SIGNAL = 2.4
ATTN: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
ATTN: EXPRESSED IN BONE MARROW, SIGNAL = 1.9
ATTN: EST HUMAN HIT: BE243764.1, EVALUE 7.00e-25
ATTN: SWISSPROT HIT: F42701, EVALUE 4.00e-33
085
3.9%; Score 7; DB 9; Length 72;
ilarity 100.0%; Pred. No. 28;
Conservative 0; Mismatches 0; Indels 0; Gaps 0;
ERTEP 20
|||||
ERTEP 34

206
, Application US/09864761
0020048763A1
ATTN:
an, Sharon G.
ank, David R.
anzel, David K.
ien, Wensheng
ATTN: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
ATTN: GENE EXPRESSION ANALYSIS BY MICROARRAY
3: Aomica-X-1
ATION NUMBER: US/09/864,761
3 DATE: 2001-05-23
TION NUMBER: US 60/180,312
DATE: 2000-02-04
TION NUMBER: US 60/207,456
DATE: 2000-05-26
TION NUMBER: US 09/632,366
DATE: 2000-08-03
TION NUMBER: GB 24263.6
DATE: 2000-10-04
TION NUMBER: US 60/236,359
DATE: 2000-09-27
TION NUMBER: PCT/US01/00666
DATE: 2001-01-30
TION NUMBER: PCT/US01/00667
DATE: 2001-01-30
TION NUMBER: PCT/US01/00664
DATE: 2001-01-30
TION NUMBER: PCT/US01/00669
DATE: 2001-01-30
TION NUMBER: PCT/US01/00665
DATE: 2001-01-30
TION NUMBER: PCT/US01/00668
DATE: 2001-01-30
TION NUMBER: PCT/US01/00663
DATE: 2001-01-30
TION NUMBER: PCT/US01/00662
DATE: 2000-09-21
TION NUMBER: US 09/608,408
DATE: 2000-06-30
TION NUMBER: US 09/774,203
DATE: 2001-01-29

; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 35206
; LENGTH: 88
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO NM 006994.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.94
; OTHER INFORMATION: EST HUMAN HIT: AW591602.1, EVALUE 5.00e-46
; OTHER INFORMATION: SWISSPROT HIT: P18892, EVALUE 5.00e-19
US-09-864-761-35206

Query Match 3.9%; Score 7; DB 9; Length 88;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 41 DGGIHLE 47
Db 9 DGGIHLE 15

RESULT 9
US-09-864-761-36770
; Sequence 36770, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-29

```

09:21:14 2004

us-09-147-801d-4.Oligo.rapb

ION NUMBER: US 60/234,687
ATE: 2000-09-21
ION NUMBER: US 09/608,408
ATE: 2000-06-30
ION NUMBER: US 09/774,203
ATE: 2001-01-29
ID NOS: 49117
max Sequence Listing Engine vers. 1.1

o sapiens

TION: MAP TO AL021917.1
TION: EXPRESSED IN BT474, SIGNAL = 1.1
TION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
TION: EXPRESSED IN HELI00, SIGNAL = 1
TION: EXPRESSED IN HELA, SIGNAL = 1.1
TION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3
TION: EXPRESSED IN BRAIN, SIGNAL = 1
TION: EXPRESSED IN LUNG, SIGNAL = 2.4
TION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
TION: EXPRESSED IN HEART, SIGNAL = 1.3
TION: EST HUMAN HIT: BE814862.1, EVALUO 1.00e-44
TION: SWISSPROT HIT: P18892, EVALUO 1.00e-20
70

3.9%: Score 7; DB 9; Length 93;
larity 100.0%; Pred. No. 34; Indels 0; Gaps 0;
Conservative 0; Mismatches 0; Indels 0; Gaps 0;

IHLE 47
|||||
IHLE 20

pplication US/10372876
US20030204071A1

re, Paul A. et al.
TION: 110 Human Secreted Proteins

: PZ021P1
TION NUMBER: US/10/372,876

DATE: 2003-02-26
ION NUMBER: 09/334,595

ATE: 1999-06-17
ION NUMBER: PCT/US98/27059

ATE: 1998-12-17
ION NUMBER: 60/070,923

ATE: 1997-12-18
ION NUMBER: 60/068,007

ATE: 1997-12-18
ION NUMBER: 60/068,057

ATE: 1997-12-18
ION NUMBER: 60/068,006

ATE: 1997-12-18
ION NUMBER: 60/068,369

ATE: 1997-12-19
ION NUMBER: 60/068,367

ATE: 1997-12-19
ION NUMBER: 60/068,368

ATE: 1997-12-19
ION NUMBER: 60/068,169

ATE: 1997-12-19
ION NUMBER: 60/068,057

r Application data removed - See File Wrapper or PALM.
ID NOS: 672
ntIn Ver. 2.0

o sapiens

US-10-372-876-437

Query Match 3.9%; Score 7; DB 12; Length 135;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 41 DGGIHLE 47
|||||
Db 18 DGGIHLE 24

RESULT 11

US-10-097-065-437
; Sequence 437, Application US/10097065
; Publication No. US20030055236A1
; GENERAL INFORMATION:
; APPLICANT: Moore, Paul A. et al.
; FILE REFERENCE: PZ021P1
; CURRENT APPLICATION NUMBER: US/10/097,065
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: PCT/US98/27059
; PRIOR FILING DATE: 1998-12-17
; PRIOR APPLICATION NUMBER: 60/070,923
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,007
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,057
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,006
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,369
; PRIOR FILING DATE: 1997-12-19
; PRIOR APPLICATION NUMBER: 60/068,367
; PRIOR FILING DATE: 1997-12-19
; PRIOR APPLICATION NUMBER: 60/068,368
; PRIOR FILING DATE: 1997-12-19
; PRIOR APPLICATION NUMBER: 60/068,169
; PRIOR FILING DATE: 1997-12-19
; PRIOR APPLICATION NUMBER: 60/068,053
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,064
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,054
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,008
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,365
; PRIOR FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 437
; LENGTH: 135
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-097-065-437

Query Match 3.9%; Score 7; DB 14; Length 135;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 41 DGGIHLE 47
|||||
Db 18 DGGIHLE 24

RESULT 12

US-10-424-599-159096
; Sequence 159096, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K

09:21:14 2004

us-09-147-801d-4.Oligo.rapb

```

you Yihua
to Yongwei
TION: Soy Nucleic Acid Molecules and Other Molecules Associated With
TION: Plants and Uses Thereof for Plant Improvement
E: 38-21(53223)B
TION NUMBER: US/10/424,599
DATE: 2003-04-28
ID NOS: 285684
16

/cine max
TION: Clone ID: PAT_MRT3847_114584C.1.pep
1096

3.9%; Score 7; DB 12; Length 161;
100.0%; Pred. No. 55;
Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RLTY 147
|||||
RLTY 14

1941
L, Application US/10424599
US20040031072A1
TION:
Rosa Thomas J
valic David K
you Yihua
to Yongwei
TION: Soy Nucleic Acid Molecules and Other Molecules Associated With
TION: Plants and Uses Thereof for Plant Improvement
E: 38-21(53223)B
TION NUMBER: US/10/424,599
DATE: 2003-04-28
ID NOS: 285684
11

/cine max
TION: Clone ID: PAT_MRT3847_137112C.1.pep
1941

3.9%; Score 7; DB 12; Length 220;
100.0%; Pred. No. 73;
Conservative 0; Mismatches 0; Indels 0; Gaps 0;

SPNA 154
|||||
SPNA 134

lication US/10197844
US20030166898A1
MATION:
WT: CHOPRA, ARVIND
OLSEN, HENRIK
GENTZ, REINER
RUBIN, STEVEN M.
F INVENTION: MYELIN OLIGODENDROCYTE GLYCOPROTEIN-LIKE
OF SEQUENCES: 39
NDENCE ADDRESS:
RESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
REET: 1100 NEW YORK AVENUE, NW, SUITE 600
TY: WASHINGTON
```

```

STATE: DC
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/197,844
FILING DATE: 19-Jul-2002
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/015,963
FILING DATE: <unknown>
APPLICATION NUMBER: US 60/035,445
FILING DATE: 30-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0520001/EKS/RCM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 223 amino acids
TYPE: amino acid
STRANDEDNESS: No. US20030166898A1 Relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-10-197-844-7

Query Match 3.9%; Score 7; DB 14; Length 223;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 7; Conservative 0; Mismatches 0; Indels 0;

Qy 41 DGGIHLE 47
|||||
Db 159 DGGIHLE 165

RESULT 15
US-10-425-114-44067
; Sequence 44067, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules As
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 44067
; LENGTH: 245
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700786691_FLI.pep
US-10-425-114-44067

Query Match 3.9%; Score 7; DB 12; Length 245;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 7; Conservative 0; Mismatches 0; Indels 0;

Qy 148 NSSSPNA 154
```

09:21:14 2004

us-09-147-801d-4.Oligo.rapb

||||
SPNA 125

April 23, 2004, 14:14:06
s

09:21:15 2004

us-09-147-801d-4.Oligo.rpr

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

ein search, using sw model
pril 23, 2004, 14:04:49 ; Search time 20 Seconds
(without alignments)
860.914 Million cell updates/sec

S-09-147-801D-4
79
KDLMEIARMPVYKGBTERP.....QVRKMGNDKFMVPTTSNVW 179

LIGO
apop 60.0 , Gapext 60.0

83366 seqs, 96191526 residues

0
its satisfying chosen parameters: 283366

ngth: 0
ngth: 2000000000

Listing first 45 summaries

PIR 78:*
: PIR1:*
: PIR2:*
: PIR3:*
: PIR4:*

s the number of results predicted by chance to have a
er than or equal to the score of the result being printed,
ved by analysis of the total score distribution.

SUMMARIES

Query	Match	Length	DB	ID	Description
6.7	2180	2	A46182		polyprotein - echo
3.9	214	2	S40254		hypothetical DNA-b
3.9	215	2	T36637		probable DNA-bind
3.9	394	1	S77216		hypothetical prote
3.9	408	2	C86156		Ti4P4.9 protein -
3.9	473	2	F89931		dihydrolipoamide d
3.9	480	2	G75313		molybdate metaboli
3.9	487	2	S60675		hypothetical prote
3.9	660	2	T20228		hypothetical prote
3.9	662	2	I37892		IL12 receptor comp
3.9	859	2	T43701		DNA-directed RNA p
3.9	1188	2	T05846		DNA-directed RNA p
3.9	1191	2	S65068		DNA-directed RNA p
3.9	1193	2	E88445		protein c2686.4 [i
3.9	1345	2	B71608		DNA-directed RNA p
3.9	1679	2	T50091		yeast Ecm29 cell w
3.4	66	2	E75432		hypothetical prote
3.4	67	2	C96765		6-phosphofructo-2-
3.4	70	2	S74243		hypothetical prote
3.4	95	2	A3571		hypothetical prote
3.4	96	2	E64341		repetitive protein
3.4	118	1	RPXFX		unknown protein en
3.4	119	2	A85816		response regulator
3.4	122	2	E87622		hypothetical prote
3.4	128	2	H84223		hypothetical prote
3.4	128	2	D95181		hypothetical prote
3.4	128	2	G98048		hypothetical prote
3.4	134	2	T29690		hypothetical prote
3.4	140	2	S74550		hypothetical prote

30 6 3.4 143 2 H70148 ribosomal
31 6 3.4 145 2 S36299 T-cell rec
32 6 3.4 159 2 A70515 probable t
33 6 3.4 161 2 D87406 ribosomal
34 6 3.4 168 2 T36405 probable p
35 6 3.4 176 2 G72365 heat shock
36 6 3.4 178 1 I40072 translati
37 6 3.4 185 2 C69117 conserved
38 6 3.4 189 2 F89732 protein ZK
39 6 3.4 194 2 T11928 ribosomal
40 6 3.4 200 2 A84994 hypothetic
41 6 3.4 206 1 TVBYQ2 GTP-bindir
42 6 3.4 212 2 S74288 hypothetic
43 6 3.4 216 2 B82020 ABC transp
44 6 3.4 221 2 AH1501 hypothetic
45 6 3.4 224 2 AI3276 hypothetic

ALIGNMENTS

RESULT 1

A46182
polyprotein - echovirus 22
N;Contains: protein 2A; protein 2B; protein 2C; protein 3A; protein 3
C;Species: echovirus 22
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct
C;Accession: A46182
R;Hyppia, T.; Horsnell, C.; Maaronen, M.; Khan, M.; Kalkkinen, N.; Al
Proc. Natl. Acad. Sci. U.S.A. 89, 8847-8851, 1992
A;Title: A distinct picornavirus group identified by sequence analysi
A;Reference number: A46182; MUID:92409614; PMID:1528901
A;Accession: A46182
A;Status: preliminary
A;Molecule type: genomic RNA; protein
A;Residues: 1-2180 <HY>
A;Cross-references: GB:S45208; GB:I00675; NID:9256078; PIDN:AB23363.
A;Experimental source: strain Harzris
A;Note: sequence extracted from NCBI backbone (NCBIN:114262, NCBIP:1;
C;Keywords: polyprotein

Query Match 6.7%; Score 12; DB 2; Length 2180;
Best Local Similarity 100.0%; Pred. No. 0.00025;
Matches 12; Conservative 0; Mismatches 0; Indels 0;

Oy 141 VLNRLTYNSSSP 152
Db 503 VLNRLTYNSSSP 514

RESULT 2

S40254
hypothetical DNA-binding protein - Streptomyces lactamdurans
C;Species: Streptomyces lactamdurans
C;Date: 19-May-1994 #sequence_revision 26-May-1995 #text_change 22-Oct
C;Accession: S40254
R;Coque, J.
submitted to the EMBL Data Library, February 1993
A;Reference number: S40253
A;Accession: S40254
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-214 <CO>
A;Cross-references: EMBL:221683; NID:9438195; PIDN:CAA79799.1; PID:9;
A;Note: the source is designated as Nocardia lactamdurans

Query Match 3.9%; Score 7; DB 2; Length 214;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0;

Oy 17 RTEPGGT 23
Db 50 RTEPGGT 56

ding protein - Streptomyces coelicolor
 tomyces coelicolor
 999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 637
 rris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 EMBL Data Library, June 1999
 er: Z21610
 537
 inary; translated from GB/EMBL/DDBJ
 DNA
 5 <OLI>
 es: EMBL:AL078610; PIDN:CAB44405.1; GSPDB:GN00070; SCOEDB:SCH35.22c
 source: strain A3(2)
 CH35.22c

3.9%; Score 7; DB 2; Length 215;
 ilarity 100.0%; Pred. No. 11;
 Conservative 0; Mismatches 0; Indels 0; Gaps 0;

EPGGT 23
 |||||
 EPGGT 57

tein sll1358 - Synechocystis sp. (strain PCC 6803)
 hocystis sp.
 803
 999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
 216
 to, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
 S.; Shimpou, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
 136, 1996
 e analysis of the genome of the unicellular cyanobacterium Synechocystis

er: S74322; MUID:97061201; PMID:8905231
 216
 c acid sequence not shown; translation not shown
 DNA
 '4 <KAN>
 es: EMBL:D90907; GB:AB001339; NID:gl652618; PIDN:BAAL17550.1; PID:gl65263
 eotide sequence was submitted to the EMBL Data Library, June 1996
 onerved hypothetical protein yoaN

3.9%; Score 7; DB 1; Length 394;
 ilarity 100.0%; Pred. No. 19;
 Conservative 0; Mismatches 0; Indels 0; Gaps 0;

TGWLRL 138
 |||||
 TGWLRL 10

- Arabidopsis thaliana
 .dopsis thaliana (mouse-ear cress)
 2001 #sequence_revision 02-Mar-2001 #text_change 27-Nov-2001
 156
 : Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 ing, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 hes, B.; Huizar, L.
 -820, 2000
 er, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
 y, T.; Rowley, D.; Sakano, H.
 erg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidop
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: C86156
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-408 <STO>
 A:Cross-references: GB:AE005172; NID:g9972375; PIDN:AAG10625.1; GSPD
 C:Genetics:
 A:Map position: 1

Query Match 3.9%; Score 7; DB 2; Length 408;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 131 GSVIGML 137
 |||||
 Db 399 GSVIGML 405

RESULT 6
 F89931
 dihydroliipoamide dehydrogenase [imported] - Staphylococcus aureus (e
 C:Species: Staphylococcus aureus
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-C
 C:Accession: F89931
 R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayast
 ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaitc
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu,
 Lancet 357, 1225-1240, 2001
 A:Title: Whole genome sequencing of methicillin-resistant Staphylococc
 A:Reference number: A89758; MUID:21311952; PMID:11418146
 A:Accession: F89931
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-473 <KUR>
 A:Cross-references: GB:BA000018; PID:gl3701317; PIDN:BAB42611.1; GSE
 A:Experimental source: strain N315
 C:Genetics:
 A:Gene: SA1349
 C:Superfamily: dihydroliipoamide dehydrogenase; dihydroliipoamide dehy
 C:Keywords: redox-active disulfide
 F;43-48/Disulfide bonds: redox-active #status predicted

Query Match 3.9%; Score 7; DB 2; Length 473;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 107 DIGLNNT 113
 |||||
 Db 283 DIGLNNT 289

RESULT 7
 G75313
 molybdate metabolism regulator-related protein - Deinococcus radiod
 C:Species: Deinococcus radiodurans
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-
 C:Accession: G75313
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson,
 , M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback,
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus
 A:Reference number: A75250; MUID:20036896; PMID:10567266
 A:Accession: G75313
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-480 <WHI>
 A:Cross-references: GB:AE002046; GB:AE000513; NID:g6459901; PIDN:AAI
 A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DR2108

3.9%; Score 7; DB 2; Length 480;
 larity 100.0%; Pred. No. 22;
 Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RGSV 133
 |||||
 RGSV 271

ein C - Corynebacterium glutamicum plasmid pGAL
 bacterium glutamicum
 97 #sequence_revision 25-Apr-1997 #text_change 22-Oct-1999
 75
 tek, M.; Hochmannova, J.; Abrahamova, Z.
 EMBL Data Library, August 1995
 mplete nucleotide sequence of the cryptic plasmid pGAL from Corynebacte
 r: S60673
 75
 nary
 DNA
 ' <NES>
 s: EMBL:X90817; NID:g951006; PIDN:CAA62330.1; PID:g951009
 i pGAL

3.9%; Score 7; DB 2; Length 487;
 larity 100.0%; Pred. No. 23;
 Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1E1AR 8
 |||||
 1E1AR 229

ein C54G10.3 - Caenorhabditis elegans
 habditis elegans
 99 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
 28
 EMBL Data Library, June 1996
 r: Z19240
 28
 nary; translated from GB/EMBL/DBJ
 DNA
 ' <WIL>
 s: EMBL:Z75532; PIDN:CAA99810.1; GSPDB:GN00023; CESP:C54G10.3
 urce: Clone C54G10

110.3
 133/2; 305/3; 336/2; 442/1; 628/3
 assigned ATP-binding cassette proteins; ATP-binding cassette homology

3.9%; Score 7; DB 2; Length 660;
 larity 100.0%; Pred. No. 30;
 Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1LKLT 71
 |||||
 1LKLT 324

ponent - human
 sapiens (man)
 96 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999

C;Accession: I37892
 R;Chua, A.O.; Chizzonite, R.; Desai, B.B.; Truitt, T.P.; Nunes, P.; I
 J;Immunol. 153, 128-136, 1994
 A;Title: Expression cloning of a human IL-12 receptor component. A n
 A;Reference number: I37892; MUID:94267217; PMID:7911493
 A;Accession: I37892
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-662 <RES>
 A;Cross-references: EMBL:U03187; NID:g507150; PIDN:AAA21340.1; PID:g

Query Match 3.9%; Score 7; DB 2; Length 662;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 7; Conservative 0; Mismatches 0; Indels 0;

Qy 14 KGERTEP 20
 |||||
 Db 625 KGERTEP 631

RESULT 11
 T43701
 DNA-directed RNA polymerase (EC 2.7.7.6) II largest chain [imported]
 C;Species: Caenorhabditis elegans
 C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 02-S
 C;Accession: T43701
 R;Sidow, A.; Thomas, W.K.
 Curr. Biol. 4, 596-603, 1994
 A;Title: A molecular evolutionary framework for eukaryotic model org
 A;Reference number: Z22636; MUID:95041334; PMID:7953533
 A;Accession: T43701
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-859 <SID>
 A;Cross-references: EMBL:U10333; NID:g520506; PIDN:AAA50224.1; PID:g
 C;Superfamily: DNA-directed RNA polymerase 132K polypeptide
 C;Keywords: nucleotidyltransferase

Query Match 3.9%; Score 7; DB 2; Length 859;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 7; Conservative 0; Mismatches 0; Indels 0;

Qy 141 VLNRLTY 147
 |||||
 Db 276 VLNRLTY 282

RESULT 12
 T05846
 DNA-directed RNA polymerase (EC 2.7.7.6) II second largest chain - A
 N;Alternate names: protein F17L22.170
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 09-Apr-1999 #sequence_revision 09-Apr-1999 #text_change 20-S
 C;Accession: T05846; S30229; S30228
 R;Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J
 submitted to the Protein Sequence Database, February 1999
 A;Reference number: Z15454
 A;Accession: T05846
 A;Molecule type: DNA
 A;Residues: 1-1188 <BEV>
 A;Cross-references: EMBL:AL035527; GSPDB:GN00062; ATSP:F17L22.170
 A;Experimental source: cultivar Columbia; BAC clone F17L22
 R;Larkin, R.; Guilfoyle, T.
 Nucleic Acids Res. 21, 1038, 1993
 A;Title: The second largest subunit of RNA polymerase II from Arabid
 A;Reference number: S30228; MUID:93197129; PMID:8451172
 A;Accession: S30229
 A;Status: nucleic acid sequence not shown
 A;Molecule type: DNA
 A;Residues: 1-353 'LY'.356-1188 <LAR>
 A;Cross-references: EMBL:Z19121; NID:gl6485; PIDN:CAA79528.1; PID:gl
 A;Accession: S30228
 A;Status: nucleic acid sequence not shown; translation not shown

mRNA
3, 'LY', 356-786, 'N', 788-1188 <LAW>
es: EMBL:Z19120; NID:G16487; PIDN:CAA79527.1; PID:G16488
eotide sequence was submitted to the EMBL Data Library, December 1992

L22.170
4
157/3; 254/3; 320/3; 342/2; 371/2; 411/3; 426/3; 472/3; 497/1; 522/3; 5
NA-directed RNA polymerase 132K polypeptide
binding; nucleotidyltransferase; nucleus; transcription
3.9%; Score 7; DB 2; Length 1188;
ilarity 100.0%; Pred. No. 49;
Conservative 0; Mismatches 0; Indels 0; Gaps 0;
NRLTY 147
|||||
NRLTY 479

polymerase (EC 2.7.7.6) II second largest chain - tomato
ersicon esculentum (tomato)
996 #sequence_revision 13-Mar-1997 #text_change 18-Jun-1999
068
Symons, R.H.
30, 337-342, 1996
e analysis of the second largest subunit of tomato RNA polymerase II.
er: S65068; MUID:96178872; PMID:8616257
068
c acid sequence not shown
mRNA
91 <WAR>
es: EMBL:U28403; NID:G1049067; PIDN:AAC49273.1; PID:G1049068

NA-directed RNA polymerase 132K polypeptide
binding; nucleotidyltransferase; nucleus; transcription
3.9%; Score 7; DB 2; Length 1191;
ilarity 100.0%; Pred. No. 49;
Conservative 0; Mismatches 0; Indels 0; Gaps 0;
NRLTY 147
|||||
NRLTY 478

[imported] - Caenorhabditis elegans
rhabditis elegans
001 #sequence_revision 10-May-2001 #text_change 24-May-2001
445
C. elegans Sequencing Consortium.
2-2018, 1998
sequence of the nematode C. elegans: a platform for investigating biolog
er: A75000; MUID:99069613; PMID:9851916
ites genome.wat1.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_ele
d errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
1445
inary
DNA
93 <STO>
es: GB:chr_III; PIDN:AAA21158.1; PID:G532805; GSPDB:GN00021; CESP:C266E.

3
NA-directed RNA polymerase 132K polypeptide
3.9%; Score 7; DB 2; Length 1193;
ilarity 100.0%; Pred. No. 49;

Matches 7; Conservative 0; Mismatches 0; Indels 0;
QY 141 VLNRLTY 147
|||||
Db 475 VLNRLTY 481

RESULT 15
B71608
DNA-directed RNA polymerase (EC 2.7.7.6) II second largest chain - m
N/Alternate names: protein PFB0715W
C/Species: Plasmodium falciparum
C/Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 20-S
C/Accession: B71608
R/Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravin
; Perte, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; Wh
Science 282, 1126-1132, 1998
A/Title: Chromosome 2 sequence of the human malaria parasite Plasmod
A/Reference number: A71600; MUID:99021743; PMID:9804551
A/Accession: B71608
A/Status: preliminary; nucleic acid sequence not shown; translation
A/Molecule type: DNA
A/Residues: 1-1345 <GAR>
A/Cross-references: GB:AE001413; GB:AE001362; NID:G3845255; PIDN:AAC
A/Experimental source: clone 3D7
C/Genetics:
A/Gene: PFB0715W
C/Superfamily: DNA-directed RNA polymerase 132K polypeptide
C/Keywords: DNA binding; nucleotidyltransferase; nucleus; transcript

Query Match 3.9%; Score 7; DB 2; Length 1345;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0;
QY 141 VLNRLTY 147
|||||
Db 493 VLNRLTY 499

Search completed: April 23, 2004, 14:08:49
Job time : 22 secs

09:21:16 2004

us-09-147-801d-4.Oligo.rsp

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

tein search, using sw model

April 23, 2004, 14:01:09 ; Search time 18 Seconds
(without alignments)
517.809 Million cell updates/sec

US-09-147-801D-4

1 KDLMEIARMPVYKGRTEP.....QVKMGNDAKFMVPTTSNIW 179

OLIGO

Gapop 60.0 , Gapext 60.0

141681 seqs, 52070155 residues

0

hits satisfying chosen parameters: 141681

length: 0
length: 2000000000

Listing first 45 summaries

Swissprot_42:*

is the number of results predicted by chance to have a
ter than or equal to the score of the result being printed,
ived by analysis of the total score distribution.

SUMMARIES

Query	Match	Length	DB	ID	Description
6.7	2180	1	POLG_EC22H	Q66578	e genome po
5.6	2188	1	POLG_EC23C	Q9yid8	e genome po
4.5	2179	1	POLG_EC23W	Q73556	e genome po
3.9	215	1	VH22_MXVL	Q9q85	myxoma viru
3.9	524	1	C6G1_DROME	Q9v674	drosohila
3.9	662	1	I12R_HUMAN	P42701	homo sapien
3.9	1188	1	RPB2_ARATH	P38420	arabidopsis
3.9	1191	1	RPB2_LYCES	Q42877	lycopersico
3.9	1193	1	RPB2_CABEL	Q10578	caenorhabdi
3.4	70	1	F261_MOUSE	P70266	mus musculu
3.4	96	1	Y333_METJA	Q57779	methanococ
3.4	137	1	CST5_MOUSE	Q9z0h6	mus musculu
3.4	143	1	RL11_BORBU	O51354	borrelia bu
3.4	159	1	BFR_MYCTU	O08465	mycobacteri
3.4	176	1	HSLV_THEMEA	Q9wy21	thermotoga
3.4	179	1	IF3_BUCAP	P46243	buchnera ap
3.4	181	1	IF3_BUCBP	P59446	buchnera ap
3.4	183	1	IF3_VIRCH	O68844	vibrio chol
3.4	188	1	Y4XK_RHISN	P55703	rhizobium s
3.4	194	1	RM06_PROWI	P46748	prototheca
3.4	200	1	Y549_BUCAI	P57614	buchnera ap
3.4	206	1	YPT1_YEAST	P01123	saccharomyc
3.4	208	1	YK69_CABEL	P34347	caenorhabdi
3.4	232	1	RL1_XANAC	Q8pnt3	xanthomonas
3.4	232	1	RL1_XANCP	Q8rtj4	xanthomonas
3.4	234	1	VGP8_EBV	P03224	epstein-bar
3.4	244	1	IOD1_CANFA	P49894	canis famil
3.4	245	1	KAD_CHLTR	O84130	chlamydia t
3.4	248	1	SODF_SOYBN	P28759	glycine max
3.4	250	1	LUXR_VIBFI	P12746	vibrio fisc
3.4	250	1	LUXS_VIBFI	P35327	vibrio fisc
3.4	253	1	KAD_CHLMU	Q9pkro	chlamydia m
3.4	258	1	ATE_BRAVA	Q89k72	bradyrhizob

34	6	3.4	260	1	NADE_HELPJ	Q9zmb0	he
35	6	3.4	260	1	NADE_HELPY	O25096	he
36	6	3.4	263	1	LPXA_CAUCR	Q9a715	ce
37	6	3.4	268	1	MAE1_HUMAN	Q15691	hc
38	6	3.4	279	1	YQCD_HAEIN	P44153	ha
39	6	3.4	292	1	ARGB_METMA	Q8pxj8	me
40	6	3.4	297	1	GALF_SALTY	P26390	sa
41	6	3.4	298	1	GALF_KLEPN	Q48447	kl
42	6	3.4	299	1	ARGB_METAC	Q8thj9	me
43	6	3.4	307	1	PRB2_CORGL	Q8ns12	cc
44	6	3.4	307	1	YTR7_ANASP	O8y8us	ar
45	6	3.4	327	1	FAH2_BACCR	Q81f42	be

ALIGNMENTS

RESULT 1
POLG_EC22H STANDARD; PRT; 2180 AA.
AC Q66578; Q90062;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polypeptide [Contains: Coat protein VP0 (P1AB); Coat pr
DE VP3 (P1C); Coat protein VP1 (P1D); Core protein P2A; Core prote
DE Core protein P2C; Core protein P3A; Genome-linked protein VPG (f
DE Picornain 3C (EC 3.4.22.28) (Protease 3C) (P3C); RNA-directed f
DE Polymerase (EC 2.7.7.48) (P3D)].
OS Echovirus 22 (strain Harris) (Human parechovirus 1).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornavi
OC Parechovirus.
OX NCBI_TaxID=103911;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 29-61; 67-80; 290-297 AND 5
RX MEDLINE=92409614; PubMed=1528901;
RA Hyypia T., Horsnell C., Maaronen M., Khan M., Kalkkinen N.,
RA Auvinen P., Kinnunen L., Stanway G.;
RA "A distinct picornavirus group identified by sequence analysis.
RL Proc. Natl. Acad. Sci. U.S.A. 89:8847-8851(1992).
RN [2]
RP SEQUENCE OF 6-22; 29-61; 67-80 AND 118-131.
RX MEDLINE=95056053; PubMed=7966616;
RA Stanway G., Kalkkinen N., Roivainen M., Ghazi F., Khan M., Smyt
RA Meurman O., Hyypia T.;
RT "Molecular and biological characteristics of echovirus 22, a
RT representative of a new picornavirus group.";
RL J. Virol. 68:8232-8238(1994).
CC -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES
AT CERTAIN Q/G SITES IN THE POLYPROTEIN. IT IS A CYSTEINE
PROTEASE.
CC -!- CATALYTIC ACTIVITY: Selective cleavage of Gln-Gly bond ir
CC poliovirus polyprotein. In other picornavirus reactions Gln
CC substituted for Gln, and Ser or Thr for Gly.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphospha
CC (RNA) (N).
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNI
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP0,
CC AND VP3.
CC -!- PTM: Specific enzymatic cleavages in vivo yield mature prot
CC ALL CLEAVAGES ARE CATALYZED BY P3C.
CC -!- PTM: THE N-TERMINUS OF VP0 IS BLOCKED.
CC -!- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

CC This SWISS-PROT entry is copyrighted. It is produced through a cc
CC between the Swiss Institute of Bioinformatics and the EMBL c
CC the European Bioinformatics Institute. There are no restricti
CC use by non-profit institutions as long as its content is
CC modified and this statement is not removed. Usage by and for
CC entities requires a license agreement (See <http://www.isb-sib.ch>
CC or send an email to license@isb-sib.ch).

CC EMBL; L02971; AAA72291.1; -.


```

OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirir
OC Leporipoxvirus.
OC NCBI_TaxID=31530;
OC [1]
OC RN
OC RP SEQUENCE FROM N.A.
OC RX MEDLINE=20032073; PubMed=10562494;
OC RA Cameron C., Hota-Mitchell S., Chen L., Barrett J., Cao J.X.,
OC RA Macaulay C., Miller D., Evans D., McFadden G.;
OC RT "The complete DNA sequence of myxoma virus.";
OC RL Virology 264:298-318(1999).
OC CC -!- FUNCTION: Plays a role for multiplication of the virus in
OC CC different cell types (By similarity).
OC CC -!- SIMILARITY: Belongs to the poxviruses C7 family.
OC CC -----
OC CC This SWISS-PROT entry is copyright. It is produced through a cc
OC CC between the Swiss Institute of Bioinformatics and the EMBL c
OC CC the European Bioinformatics Institute. There are no restricti
OC CC use by non-profit institutions as long as its content is
OC CC modified and this statement is not removed. Usage by and for
OC CC entities requires a license agreement (See http://www.isb-sib.c
OC CC or send an email to license@isb-sib.ch).
OC CC -----
OC CC EMBL; AF170726; AAF14951.1; -
OC CC InterPro; IPR004967; Pox C7 F8A.
OC CC Pfam; PF03287; Pox C7 F8A; 1.
OC CC PIRSF; PIRSF003779; VAC_C7L1; 1.
OC CC SEQUENCE 215 AA; 24578 MW; 9D5D3CF8FCB74B27 CRC64;
OC CC -----
OC CC
OC DR Query Match 3.9%; Score 7; DB 1; Length 215;
OC DR Best Local similarity 100.0%; Pred.No.4.8;
OC DR Matches 7; Conservative 0; Mismatches 0; Indels 0;
OC CC -----
OC QY 78 NKGRLRM 84
OC DB 153 NKGRLRM 159
OC DB |||||
OC DB -----
OC RESULT 5
OC ID CGGI DROME STANDARD; PRT; 524 AA.
OC AC Q9V674; O76800; Q95S18;
OC DT 16-OCT-2001 (Rel. 40, Created)
OC DT 16-OCT-2001 (Rel. 40, Last sequence update)
OC DT 10-OCT-2003 (Rel. 42, Last annotation update)
OC DE Cytochrome P450 6G1 (EC 1.14.-.-) (CYP6G1) (Cyp6-like protein)
OC DE CYP6G1 OR CYP6-LIKE OR RST(2)DPT OR DDT-R OR CG8453.
OC GS Drosophila melanogaster (Fruit fly).
OC OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC OC Ephydroidea; Drosophilidae; Drosophila.
OC ON NCBI_TaxID=7227;
OC ON [1]
OC ON SEQUENCE FROM N.A.
OC RP TISSUE=Larva;
OC RC Danielson P.B., Al-Zahrani A., Fogleman J.C.;
OC RA "Isolation of a novel CYP6-like cytochrome P450 from Drosophila
OC RT melanogaster.";
OC RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
OC RN [2]
OC RN SEQUENCE FROM N.A., FUNCTION, AND POLYMORPHISM.
OC RP STRAIN=Canton-S, Hikone-R, and WC2;
OC RX MEDLINE=24204273; PubMed=12351787;
OC RA Daborn P.J., Yen J.L., Bogwitz M.R., Le Goff G., Feil E., Jeffe
OC RA Tijet N., Perry T., Heckel D., Batterham P., Feyereisen R.,
OC RA Wilson T.G., ffrench-Constant R.H.;
OC RT "A single P450 allele associated with insecticide resistance in
OC RT Drosophila.";
OC RN Science 297:2253-2256(2002).
OC RN [3]
OC RN SEQUENCE FROM N.A.
OC RP RA Buzeli R.A., Pedra J.H.F., Scharf M., Pittendrigh B.R.;
OC RN Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
OC RN [4]

```

OM N.A.
 96006; PubMed=10731132;
 Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 P.G., Scher S.E., Li P.W., Hoskins R.A., Calle R.F.,
 Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
 Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 Oyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 Cawley S., Dahlke C., Davenport L.B., Davies P.,
 DeLcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 ei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 tei B., McIntosh T.C., McLeod M.P., McPherson D.,
 Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 Stadler A.C., Stapleton M., Strong R., Sun E.,
 Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 Wassarman D.A., Weinstock G.M., Weissenbach J.,
 M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 Myers E.W., Rubin G.M., Venter J.C.;
 sequence of *Drosophila melanogaster*.
 :2185-2195(2000).
 216-524 FROM N.A.
 eley; TISSUE=Head;
 26066; PubMed=12537569;
 Carlson J.W., Brokstein P., Yu C., Champe M.,
 Guarin H., Krommiller B., Pacle J.M., Park S., Wan K.H.,
 Celniker S.E.;
 la full-length cDNA resource";
 3:RESEARCH0080.1-RESEARCH0080.8(2002).
 on-S. Hikone-R. and Wisconsin-1;
 69003; PubMed=1181026;
 Boundy S., Yen J., Pittendrigh B., French-Constant R.H.;
 ance in *Drosophila* correlates with Cyp6g1 over-expression
 cross-resistance to the neonicotinoid imidacloprid";
 Genomics 266:556-563(2001).
 N: Necessary and sufficient for resistance to insecticides
 imidacloprid. May be involved in the metabolism of insect
 s.
 IC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
 d flavoprotein + H(2)O.
 ULAR LOCATION: Membrane-bound. Endoplasmic reticulum
 ial).
 PHISM: Hikone-R and WC2 are insecticide resistant strains,
 S and Oregon-RC are insecticide susceptible.
 ITY: Belongs to the cytochrome P450 family.

 PROT entry is copyright. It is produced through a collaboration
 e Swiss Institute of Bioinformatics and the EMBL outstation -
 n Bioinformatics Institute. There are no restrictions on its
 n-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for
 entities requires a license agreement (See <http://www.isb-sib.ch>)
 or send an email to license@isb-sib.ch.

 CC EMBL; AF083946; AAC33298.1; -;
 CC EMBL; AF081960; AAL89788.1; -;
 CC EMBL; AF003823; AAF58557.1; -;
 CC EMBL; AY060770; AAL28318.1; ALT_INIT.
 CC HSP; PI4779; LJPZ.
 CC Flybase; Fgn0025454; Cyp6g1.
 CC GO; GO:0005789; C: endoplasmic reticulum membrane; NAS.
 CC GO; GO:0008163; P: DDT resistance; IMP.
 CC GO; GO:0017085; P: response to insecticide; IMP.
 CC InterPro; IPR001128; Cytochrome_P450.
 CC Pfam; PF00067; P450; 1.
 CC PRINTS; PR00385; P450.
 CC PROSITE; PS00086; CYTOCHROME_P450; 1.
 CC Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
 KW Endoplasmic reticulum.
 FT METAL 458 458 IRON (HEME AXIAL LIGAND) (BY SIMIL
 FT CONFLICT 31 34 KGIP -> RHT (IN REF. 1).
 FT CONFLICT 39 43 TPIIG -> RDHCNTKVFVKWIGLDC (IN RE
 FT CONFLICT 59 59 L -> P (IN REF. 1).
 FT CONFLICT 108 120 AFCDPHGDPGLYN -> GELRIPWRSIGLY (IN
 FT CONFLICT 129 133 HWKGI -> IREF (IN REF. 1).
 FT CONFLICT 153 153 M -> L (IN REF. 1).
 FT CONFLICT 274 274 D -> S (IN REF. 1).
 FT CONFLICT 304 305 AQ -> VS (IN REF. 1).
 FT CONFLICT 500 500 L -> H (IN REF. 1).
 FT CONFLICT 524 524 AA; 59866 MW; 35F41CCCA866756F CRC64;
 SQ SEQUENCE 524 AA; 59866 MW; 35F41CCCA866756F CRC64;
 Query Match 3.9%; Score 7; DB 1; Length 524;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 7; Conservative 0; Mismatches 0; Indels 0;
 QY 41 DGGIHLE 47
 DB 503 DGGIHLE 509
 RESULT 6
 IL12R HUMAN STANDARD; PRT; 662 AA.
 AC P42701;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Interleukin-12 receptor beta-1 chain precursor (IL-12R-beta1)
 DE (Interleukin-12 receptor beta) (IL-12 receptor beta component)
 DE (IL-12RB1).
 GN IL12RB1 OR IL12RB OR IL12R.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostom
 OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94267217; PubMed=7911493;
 RA China A.O., Chizzonite R., Desai B.B., Truitt T.P., Nunes P.,
 RA Minetti L.J., Warrior R.R., Presky D.H., Levine J.F., Gately M.
 RA Gubler U.;
 RT "Expression cloning of a human IL-12 receptor component. A new
 RT of the cytokine receptor superfamily with strong homology to gp
 RL J. Immunol. 153:128-136(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Elloumi-Zghal H., Abdelhak S., Dellagi K.;
 RT "Genomic structure of IL12RB1 gene."
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SUBUNITS.
 RX MEDLINE=97098510; PubMed=8943050;

Yang H., Minetti L.J., Chua A.O., Nabavi N., Wu C.-Y., Gubler U.;
Interleukin 12 receptor complex is composed of two cytokine receptor subunits.";
Acad. Sci. U.S.A. 93:14002-14007(1996).

-213.
16722; PubMed11424023;
Enser A., Breiman A., Reichenbach J., El Baghdadi J., Emile J.-F., Gaillard J.-L., Meinel E., Casanova J.-L.;
n-12 receptor beta-1 deficiency in a patient with abdominal s.";
Dis. 184:231-236(2001).
N: Involved in IL12 transduction. Binds to IL12 with a inity.
y: Dimer/oligomer; disulfide-linked. The functional high y IL12 receptor is composed of at least IL12RB1 and

ULAR LOCATION: Type I membrane protein.
TIVE PRODUCTS:
lternative splicing; Named isoforms=2;
ng;
=P42701-1; Sequence=Displayed;
ort;
=P42701-2; Sequence=VSP_001715;
No experimental confirmation available;
: Defects in IL12RB1 are a cause of susceptibility to mycobacterial and Salmonella infections in otherwise individuals.
ITY: Belongs to the type I cytokine family of receptors.
ly 2.
ITY: Contains 5 fibronectin type III domains.
PROT entry is copyright. It is produced through a collaboration a Swiss Institute of Bioinformatics and the EMBL outstation - a Bioinformatics Institute. There are no restrictions on its a-profit institutions as long as its content is in no way d this statement is not removed. Usage by and for commercial quires a license agreement (See <http://www.isb-sib.ch/announce/> email to license@isb-sib.ch).

7; AAA21340.1; ..
588; CAC10446.1; ..
589; CAC10446.1; JOINED.
590; CAC10446.1; JOINED.
591; CAC10446.1; JOINED.
592; CAC10446.1; JOINED.
593; CAC10446.1; JOINED.
594; CAC10446.1; JOINED.
595; CAC10446.1; JOINED.
596; CAC10446.1; JOINED.
597; CAC10446.1; JOINED.
598; CAC10446.1; JOINED.
599; CAC10446.1; JOINED.
700; CAC10446.1; JOINED.
701; CAC10446.1; JOINED.
; 137892.
; 5971; IL12RB1.
; ..
887; C: integral to plasma membrane; TAS.
517; F: Interleukin-12 receptor activity; TAS.
886; F: Transmembrane receptor activity; TAS.
960; P: Antimicrobial humoral response (sensu Inver. . . ; TAS.
166; P: cell surface receptor linked signal transdu. . . ; TAS.
284; P: positive regulation of cell proliferation; NAS.
PR008957; FN III-like.
PR003961; FN III.
PR003529; Hemopoetn L_F2.
060; FN3; 1.
01353; HEMATOPO. REC L_F2; 1.
rausmembrane; Glycoprotein; Signal; Alternative splicing; ymorphism.

FT	SIGNAL	1	23	POTENTIAL.
FT	CHAIN	24	662	INTERLEUKIN-12 RECEPTOR BETA-1 CH
FT	DOMAIN	24	545	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	546	570	POTENTIAL.
FT	DOMAIN	571	662	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	43	133	FIBRONECTIN TYPE-III 1.
FT	DOMAIN	143	236	FIBRONECTIN TYPE-III 2.
FT	DOMAIN	237	337	FIBRONECTIN TYPE-III 3.
FT	DOMAIN	338	444	FIBRONECTIN TYPE-III 4.
FT	DOMAIN	445	540	FIBRONECTIN TYPE-III 5.
FT	DISULFID	52	62	BY SIMILARITY.
FT	CARBOHYD	121	121	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	329	329	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	346	346	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	352	352	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	442	442	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	456	456	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	VARSPLIC	659	662	KAKM -> DR (in isoform Short).
FT	VARIANT	213	213	/FTIG=VSP_001715.
FT				R -> W (confers susceptibility to
FT				mycobacterial and Salmonella infec
FT				/FTIG=VAR_015577.
FT	VARIANT	365	365	M -> T (in dbSNP:375947).
FT				/FTIG=VAR_011986.
FT	VARIANT	378	378	G -> R (in dbSNP:401502).
FT				/FTIG=VAR_011987.
SQ	SEQUENCE	662 AA;	73108 MW;	541ADA60F62DA1EF CRC64;

Query Match 3.9%; Score 7; DB 1; Length 662;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY	14 KGERTEP 20
DB	625 KGERTEP 631

RESULT 7
RPB2 ARATH
ID RPB2 ARATH STANDARD; PRT; 1188 AA.
AC P38420; Q9SVS6;
DT 01-OCT-1994 (Rel. 30, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA-directed RNA polymerase II 135 kDa polypeptide (EC 2.7.7.6)
DE (RNA polymerase II subunit 2).
GN RPB135 OR RPB2 OR RPL40 OR AT4G21710 OR F17L22.170.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophy
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; ro
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=93197129; PubMed=8451172;
RA Larkin R., Guilfoyle T.J.;
RT "The second largest subunit of RNA polymerase II from Arabidops
thaliana".
RL Nucleic Acids Res. 21:1038-1038(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083488; PubMed=10617198;
RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.
RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA Harris B., Ansoorge W., Brandt P., Grivell L.A., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Muell
RA Kreis M., Delsen M., Puigdomenech P., Watson M., Schmidheini
RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancrof
RA Vos P., Hohnel J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche

Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
 r T., Bothe G., Ramepger U., Hilbert H., Braun M.,
 Brandt A., Peters S., van Staveren M., Dirkee W.,
 Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
 Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
 Buyshaert C., Gielen J., Villarroel R., De Clercq R.,
 M., Rogers J., Cronin A., Quail M.A., Bray-Allen S.,
 oggett J., Hall S., Kay M., Lennard N., McLeay K., Mayes R.,
 Rajandream M.A., Lyne M., Benes V., Rechmann S.,
 Bloeker H., Scharfe M., Grimm M., Loechnert T.-H.,
 Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
 Argirion A., Vitale D., Liuguori R., Piravandi E.,
 Quigley F., Schmbald G., Muendlein A., Feilber R.,
 Hiller R., Schmidt W., Lecharny A., Aubourg S.,
 Cooke R., Berger C., Monfort A., Casacuberta E.,
 Weber N., Vandenbol M., Barques M., Terol J., Torres A.,
 A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
 Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
 Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
 Murray J., Sheet P., Cordes M., Abu-Threideh J.,
 Kalicki J., Graves T., Harmon G., Edwards J.,
 Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 intley D., Fulton B., Miller N., Greco T., Kemp K.,
 Fulton L., Mardis E., Dante M., Pepin K., Hillier L.W.,
 Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
 J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
 Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
 g J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
 O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
 Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
 xia M.A., Martienssen R., McCombie W.R.;
 nd analysis of chromosome 4 of the plant Arabidopsis
 769-777(1999).
 N: DNA-dependent RNA polymerase catalyzes the transcription
 into RNA using the four ribonucleoside triphosphates as
 tes.
 IC ACTIVITY: N nucleoside triphosphate = N diphosphate +
).
): Each class of RNA polymerase is assembled from 9 to 14
 nt polypeptides. This subunit is the second largest
 nt of RNA polymerase II.
 ULAR LOCATION: Nuclear.
 ANEQUOS: Three distinct zinc-containing RNA polymerases are
 or, polymerase I: polymerase I for the ribosomal RNA
 SS and tRNA genes.
 ITY: Belongs to the RNA polymerase beta chain family.

 PROT entry is copyright. It is produced through a collaboration
 e Swiss Institute of Bioinformatics and the EMBL o
 n Bioinformatics Institute. There are no restrictions on its
 n-profit institutions as long as its content is in no way
 d this statement is not removed. Usage by and for commercial
 quires a license agreement (See <http://www.isb-sib.ch/announce/>
 email to license@isb-sib.ch).

 0: CAA79527.1; -
 1: CNA79528.1; -
 527: CAB36815.1; -
 555: CAB81278.1; -
 ; T05846.
 PR007121: RNA_pol_B.
 PR007644: RNA_pol_Rpb2_1.
 PR007642: RNA_pol_Rpb2_2.
 PR007645: RNA_pol_Rpb2_3.
 PR007646: RNA_pol_Rpb2_4.
 PR007647: RNA_pol_Rpb2_5.
 PR007120: RNA_pol_Rpb2_6.
 PR007641: RNA_pol_Rpb2_7.

DR Pfam: PF04563; RNA_pol_Rpb2_1; 1.
 DR Pfam: PF04561; RNA_pol_Rpb2_2; 1.
 DR Pfam: PF04565; RNA_pol_Rpb2_3; 1.
 DR Pfam: PF04566; RNA_pol_Rpb2_4; 1.
 DR Pfam: PF04567; RNA_pol_Rpb2_5; 1.
 DR Pfam: PF00562; RNA_pol_Rpb2_6; 1.
 DR Pfam: PF04560; RNA_pol_Rpb2_7; 1.
 DR PROSITE: PS01166; RNA_POL_BETA; 1.
 KW Transferase: DNA-directed RNA polymerase; Transcription; Zinc;
 K Zinc-finger; Nuclear protein.
 FT DOMAIN 14 20 ASP/GLU-RICH (ACIDIC).
 ZN FING 1145 1145 CA-TYPE (POTENTIAL).
 FT VARIANT 787 787 I -> N.
 FT CONFLICT 354 355 PH -> LY (IN REF. 1).
 SQ SEQUENCE 1188 AA; 135018 MW; C304E43515C2C364 CRC64;
 Query Match 3.9%; Score 7; DB 1; Length 1188;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 0; Indels 0;
 QY 141 VLNRLLTY 147
 Db 473 VLNRLLTY 479
 RESULT 8
 RPB2 LYCES
 ID RPB2 LYCES STANDARD; PRT; 1191 AA.
 AC Q42877;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE DNA-directed RNA polymerase II 135 kDa polypeptide (EC 2.7.7.6)
 DE (RNA polymerase II subunit 2).
 GN RPB2.
 OS Lycopersicon esculentum (Tomato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophy
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; as
 OC Lamiales; Solanales; Solanaceae; Solanum.
 OX NCBI_TaxID=4081;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Grosse liase;
 RX MEDLINE=96178872; PubMed=8616257;
 RA Warfallow D., Symons R.H.;
 RT "Sequence analysis of the second largest subunit of tomato RNA
 RT polymerase II.";
 RL Plant Mol. Biol. 30:337-342(1996).
 CC -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transc
 CC of DNA into RNA using the four ribonucleoside triphosphates
 CC substrates.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphospha
 CC (RNA) (N).
 CC -!- SUBUNIT: Each class of RNA polymerase is assembled from 9 t
 CC different polypeptides. This subunit is the second largest
 CC component of RNA polymerase II.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- MISCELLANEOUS: Three distinct zinc-containing RNA polymeras
 CC found in eukaryotic nuclei: polymerase I for the ribosomal
 CC precursor, polymerase II for the mRNA precursor, and polyme
 CC III for 5S and tRNA genes.
 CC -!- SIMILARITY: Belongs to the RNA polymerase beta chain family
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a co
 CC between the Swiss Institute of Bioinformatics and the EMBL o
 CC the European Bioinformatics Institute. There are no restricti
 CC use by non-profit institutions as long as its content is
 CC modified and this statement is not removed. Usage by and for
 CC entities requires a license agreement (See <http://www.isb-sib.ch>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U28403; AAC49273.1; -
 DR PIR: S65068; S65068.

PR007121; RNA_pol_B.
 PR007644; RNA_pol_Rpb2_1.
 PR007642; RNA_pol_Rpb2_2.
 PR007645; RNA_pol_Rpb2_3.
 PR007646; RNA_pol_Rpb2_4.
 PR007647; RNA_pol_Rpb2_5.
 PR007120; RNA_pol_Rpb2_6.
 PR007641; RNA_pol_Rpb2_7.
 53; RNA_pol_Rpb2_1; 1.
 51; RNA_pol_Rpb2_2; 1.
 55; RNA_pol_Rpb2_3; 1.
 56; RNA_pol_Rpb2_4; 1.
 57; RNA_pol_Rpb2_5; 1.
 62; RNA_pol_Rpb2_6; 1.
 60; RNA_pol_Rpb2_7; 1.
 01166; RNA_POL_BETA; 1.
 ; DNA-directed RNA polymerase; Transcription; Zinc;
 ; Nuclear protein.
 14 19 ASP/GLU-RICH (ACIDIC).
 1123 1144 C4-TYPE (POTENTIAL).
 1191 AA; 135063 MW; 8F177640C072BCD2 CRC64;

3.9%; Score 7; DB 1; Length 1191;

ilarity 100.0%; Pred.No.24;
Conservative 0; Mismatches 0; Indels 0; Gaps 0;

NRLTY 147
 |||||
 NRLTY 478

STANDARD; PRT; 1193 AA.

(Rel. 34, Created)

(Rel. 35, Last sequence update)

(Rel. 38, Last annotation update)

RNA polymerase II second largest subunit (EC 2.7.7.6)
case I subunit 2).

is elegans.

Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

Peloderinae; Caenorhabditis.

5239;

OM N.A.

col N2;

NOV-1994) to the EMBL/GenBank/DBJ databases.

200-1058 FROM N.A.

11334; PubMed=7953533;

omas W.K.;

evolutionary framework for eukaryotic model organisms.";

4:596-603 (1994).

RNA polymerase II catalyzes the transcription
into RNA using the four ribonucleoside triphosphates as
es.

C ACTIVITY: N nucleoside triphosphate = N diphosphate +
es.

Each class of RNA polymerase is assembled from 9 to 14
it polypeptides. This subunit is the second largest
it of RNA polymerase II.

JLAR LOCATION: Nuclear.

ANOUS: Three distinct zinc-containing RNA polymerases are

RNA polymerase I for the ribosomal RNA

5S and 5.8S genes.

(TV: Belongs to the RNA polymerase beta chain family.

PROT entry is copyright. It is produced through a collaboration

Swiss Institute of Bioinformatics and the EMBL outstation

the European Bioinformatics Institute. There are no restricti
 use by non-profit institutions as long as its content is
 modified and this statement is not removed. Usage by and for
 entities requires a license agreement (See <http://www.isb-sib.ch>
 or send an email to license@isb-sib.ch).

EMBL; U13875; AAA21158.1; -.

EMBL; U10333; AAA50224.1; -.

PIR; E88445; E88445.

PIR; T43701; T43701.

WormPep; C26E6.4; CE01162.

InterPro; IPR007121; RNA_pol_B.

InterPro; IPR007644; RNA_pol_Rpb2_1.

InterPro; IPR007642; RNA_pol_Rpb2_2.

InterPro; IPR007645; RNA_pol_Rpb2_3.

InterPro; IPR007646; RNA_pol_Rpb2_4.

InterPro; IPR007647; RNA_pol_Rpb2_5.

InterPro; IPR007120; RNA_pol_Rpb2_6.

InterPro; IPR007641; RNA_pol_Rpb2_7.

Pfam; PF04563; RNA_pol_Rpb2_1; 1.

Pfam; PF04561; RNA_pol_Rpb2_2; 1.

Pfam; PF04565; RNA_pol_Rpb2_3; 1.

Pfam; PF04566; RNA_pol_Rpb2_4; 1.

Pfam; PF04567; RNA_pol_Rpb2_5; 1.

Pfam; PF04562; RNA_pol_Rpb2_6; 1.

Pfam; PF04560; RNA_pol_Rpb2_7; 1.

PROSITE; PS01166; RNA_POL_BETA; 1.

Transferase; DNA-directed RNA polymerase; Transcription; Zinc;

Zinc-finger; Nuclear protein.

ZN FING 1125 1146 C4-TYPE (POTENTIAL).

SEQUENCE 1193 AA; 134904 MW; B8A85E74E9CC7BEE CRC64;

Query Match

Best Local Similarity 3.9%; Score 7; DB 1; Length 1193;

Matches 7; Conservative 100.0%; Pred.No.24; Mismatches 0; Indels 0;

QY 141 VLNRLTY 147

DB 475 VLNRLTY 481

RESULT 10

ID_F261 MOUSE

AC P70266; STANDARD; PRT; 70 AA.

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 1 (6PF-2-K/

2,6-P2ASE liver isozyme) [Includes: 6-phosphofructo-2-kinase

(EC 2.7.1.105); Fructose-2,6-bisphosphatase (EC 3.1.3.46)] (Fra

GN PFKEBI.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostom

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; M

OC NCBI TaxID=10090;

RP [1] SEQUENCE FROM N.A.

RC STRAIN=BALB/c; TISSUE=Liver;

RA MEDLINE=9640301; PubMed=8814283;

RA Batra R.S., Brown R.M., Brown G.K., Craig I.W.;

"Molecular cloning and tissue-specific expression of mouse kidn

RT phosphofructo-2-kinase/fructose-2,6-bisphosphatase.";

RL FEBS Lett. 393:167-173 (1996).

CC -! FUNCTION: Synthesis and degradation of fructose 2,6-bisphos

CC -! CATALYTIC ACTIVITY: ATP + D-fructose 6-phosphate = ADP + D-

fructose 2,6-bisphosphate.

CC -! CATALYTIC ACTIVITY: D-fructose 2,6-bisphosphate + H(2)O = D-

fructose 6-phosphate + phosphate.

CC -! ENZYME REGULATION: Phosphorylation results in inhibition of

kinase activity (By similarity).

CC -! SUBUNIT: Homodimer (By similarity).

CC -! TISSUE SPECIFICITY: Liver.

ITY: In the C-terminal section; belongs to the glycerate mutase family.

PROT entry is copyright. It is produced through a collaboration with the Swiss Institute of Bioinformatics and the EMBL outstation - in Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> email to license@isb-sib.ch).

8; CAA67353.1; --
; S74243.

3; IFBT.

PR001345; PG/BPGM_mutase.

00; PGAM; 1.

00175; PG MUTASE; PARTIAL.

onal enzyme; Transferase; Kinase; Hydrolase; ATP-binding;

tion; Liver; Multigene family.

1

1

<1 70 FRUCTOSE-2,6-BISPHOSPHATASE.

70 AA; 7963 MW; 4FC4D1269A98FC50 CRC64;

3.4%; Score 6; DB 1; Length 70;

ilarity 100.0%; Pred. No. 22;

Conservative 0; Mismatches 0; Indels 0; Gaps 0;

LKLT 71

||||

LKLT 25

STANDARD; PRT; 96 AA.

(Rel. 35, Created)

(Rel. 35, Last sequence update)

(Rel. 40, Last annotation update)

1 protein M30333.

us jannaschi.

ryarchaeota; Methanococci; Methanococcales;

ococaceae; Methanocaldococcus.

2190;

OM N.A.

1 / DSM 2661 / ATCC 43067;

37999; PubMed-8688087;

White O., Olsen G.J., Zhou L., Fleischmann R.D.,

Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,

.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,

Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,

Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,

.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,

, Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,

, Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;

genome sequence of the methanogenic archaeon, Methanococcus

1058-1073 (1996).

PROT entry is copyright. It is produced through a collaboration with the Swiss Institute of Bioinformatics and the EMBL outstation - in Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> email to license@isb-sib.ch).

7; AAB98321.1; --

; E64341.

3; --

KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 96 AA; 10927 MW; A4EB1A265B52EB99 CRC64;

Query Match 3.4%; Score 6; DB 1; Length 96;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 6; Conservative 0; Mismatches 0; Indels 0;

QY 65 STVLKL 70

|||||

Db 77 STVLKL 82

RESULT 12

CST9_MOUSE

ID CST9_MOUSE STANDARD; PRT; 137 AA.

AC 09Z0H6;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Cystatin 9 precursor (Testatin).

GN CST9 OR CRESP.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; N

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Testis;

RX MEDLINE=99045663; PubMed=9826679;

RA Toehoenen V., Oesterlund C., Nordqvist K.;

RT "Testatin: a cystatin-related gene expressed during early testi

development.";

RL Proc. Natl. Acad. Sci. U.S.A. 95:14208-14213 (1998).

RN [2]

RP SEQUENCE FROM N.A.

RA Kanno Y., Tamura M., Nakatani N.;

RT "Identification of Cresp, a novel cystatin-related gene.";

RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.

CC -! FUNCTION: Could play an essential role in tissue reorganiza

CC during early testis development.

CC -! SUBCELLULAR LOCATION: Secreted (Potential).

CC -! TISSUE SPECIFICITY: Expression is restricted to fetal gonad

CC adult testis.

CC -! SIMILARITY: Belongs to the cystatin family.

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

NTNF 114
 ||||
 NTNF 114

STANDARD; PRT; 143 AA.
 (Rel. 37, Created)
 (Rel. 37, Last sequence update)
 (Rel. 43, Last annotation update)
 il protein L11.
 }93.
 }gorferi (Lyme disease spirochete).
 }irochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
 }39;
 XM N.A.
 }35210 / B31;
 }5943; PubMed=9403685;
 }White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
 }Tomb J.-F., Fleischmann R.D., Richardson D., Hanson M.,
 }Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
 }Palmer N., Adams M.D., Cocayne J.D., Weidman J.,
 }Watthey L., McDonald L., Artiach P., Bowman C.,
 }Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
 }Venter J.C.;
 }nence of a Lyme disease spirochaete, Borrelia
 };
 }80-586(1997).
 }i: This protein binds directly to 23S ribosomal RNA (By
 }ty).
 }or more lysine residues are methylated (By similarity).
 }TY: Belongs to the L11P family of ribosomal proteins.
 }-----
 }ROT entry is copyright. It is produced through a collaboration
 }i Swiss Institute of Bioinformatics and the EMBL outstation -
 }i Bioinformatics Institute. There are no restrictions on its
 }i-profit institutions as long as its content is in no way
 }i this statement is not removed. Usage by and for commercial
 }iquires a license agreement (See <http://www.isb-sib.ch/announce/>
 }i mail to license@isb-sib.ch).
 }-----
 }45; AAC66772.1; -
 }H70148.
 }1QA6.
 }-----
 }736; -; 1.
 }R000911; Ribosomal L11.
 }R006519; Ribosomal L11bac.
 }8; Ribosomal L11; 1.
 }6; Ribosomal_L11.N; 1.
 }1367; Ribosomal_L11; 1.
 }49; RL11; 1.
 }GR01632; L11_bact; 1.
 }0359; RIBOSOMAL L11; 1.
 }rotein; rRNA-binding; Methylation; Complete proteome.
 }43 AA; 15168 MW; 0B3E933DB408A4A2 CRC64;
 }-----
 }3.4%; Score 6; DB 1; Length 143;
 }larity 100.0%; Pred.No. 43;
 }Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 }-----
 }STAR 8
 }-----
 }STAR 112

AC 008465;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Bacterioferritin (BFR)
 GN BFR OR BFR1 OR RV1876 OR MT1925 OR MTCY180.42C OR MB1907.
 OS Mycobacterium tuberculosis, and
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773, 1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Har
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekoa F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holro
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=22208494; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikul
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical
 RT laboratory strains.";
 RL J. Bacteriol. 184:5479-5490(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.bovis; STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon
 RA Harris B., Aukin R., Doggett J., Mayes R., Keating L., Wheeler
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G
 RT "The complete genome sequence of Mycobacterium bovis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 CC -!- FUNCTION: May perform analogous functions in iron detoxifica
 CC and storage to that of animal ferritins (By similarity).
 CC -!- COFACTOR: Binds 1 heme B (iron-protoporphyrin IX) group per
 CC (By similarity).
 CC -!- SUBUNIT: Oligomer of 24 identical subunits (By similarity).
 CC -!- MISCELLANEOUS: The di-iron binding site functions as active
 CC where iron ions are oxidized from iron(II) to iron(III) bef
 CC they are stored (By similarity).
 CC -!- SIMILARITY: Belongs to the bacterioferritin family.
 CC -!- SIMILARITY: Contains 1 ferritin-like diiron domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a co
 CC the Swiss Institute of Bioinformatics and the EMBL on
 CC the European Bioinformatics Institute. There are no restrictio
 CC use by non-profit institutions as long as its content is
 CC modified and this statement is not removed. Usage by and for
 CC entities requires a license agreement (See <http://www.isb-sib.ch/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z97193; CAB10050.1; -
 CC EMBL; AE007049; AAK46197.1; -
 CC EMBL; BX248340; CAD94610.1; -
 CC F07; A70515; A70515.
 CC HSP; P11056; IBCF.
 CC TIGR; MT1925; -
 CC -----
 DR EMBL; Z97193; CAB10050.1; -
 DR EMBL; AE007049; AAK46197.1; -
 DR EMBL; BX248340; CAD94610.1; -
 DR F07; A70515; A70515.
 DR HSP; P11056; IBCF.
 DR TIGR; MT1925; -

```

; RV1876; -
PR002024; Bacterioferritin.
PR008331; Ferritin Dps.
PR009040; Ferritin_like.
10: ferritin; 1.
0601; BACFERRITIN.
02289; Bacterioferritin; 1.
IGR00754; bfr; 1.
00549; BACTERIOFERRITIN; 1.
50905; FERRITIN LIKE; 1.
e; Heme; Metal-Binding; Iron; Complete proteome.
1 145
FERRITIN-LIKE DIIRON.
18 18 IRON 1 (BY SIMILARITY).
51 51 IRON 1 (BY SIMILARITY).
51 51 IRON 2 (BY SIMILARITY).
52 52 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
54 54 IRON 1 (BY SIMILARITY).
94 94 IRON 2 (BY SIMILARITY).
127 127 IRON 1 (BY SIMILARITY).
127 127 IRON 2 (BY SIMILARITY).
130 130 IRON 2 (BY SIMILARITY).
159 AA; 18341 MW; 09856D7392A9CD60 CRC64;
3.4%; Score 6; DB 1; Length 159;
ilarity 100.0%; Pred. No. 47;
Conservative 0; Mismatches 0; Indels 0; Gaps 0;
LNRL 145
|||||
LNRL 101

```

STANDARD; PRT; 176 AA.

```

(Rel. 40, Created)
(Rel. 40, Last sequence update)
(Rel. 43, Last annotation update)
nt protease hslv (EC 3.4.25.-).
521.
maritima.
thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
2336;
CM N.A.
/ DSM 3109 / ATCC 43589;
87316; PubMed-10360571;
Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
L., Smith H.O., Venter J.C., Fraser C.M.; White O.,
or lateral gene transfer between Archaea and Bacteria from
ence of Thermotoga maritima."
323-329(1999).
N; Protease subunit of a proteasome-like degradation
(BY similarity).
: A double ring-shaped homohexamer of hslv is capped on
de by a ring-shaped hslU homohexamer (By similarity).
ULAR LOCATION: Cytoplasmic (By similarity).
ITY: Belongs to peptidase family T1B. Hslv subfamily.
-----
PROT entry is copyright. It is produced through a collaboration
e Swiss Institute of Bioinformatics and the EMBL outstation -
a Bioinformatics Institute. There are no restrictions on its
a-profit institutions as long as its content is in no way
d this statement is not removed. Usage by and for commercial
quires a license agreement (See http://www.isb-sib.ch/announce/
email to license@isb-sib.ch).
-----
728; AAD35606.1; -

```

```

DR PIR; G72365; G72365.
DR HSSP; P31059; 1E94.
DR MEROPS; T01.006; -.
DR TIGR; TM0521; -.
DR HAMAP; MF_00248; -; 1.
DR InterPro; IPR001353; Peptidase_T1.
DR Pfam; PF00227; Proteasome; 1.
KW Hydrolase; Protease; Threonine protease; Complete proteome.
FT ACT_SITE 6 BY SIMILARITY.
SQ SEQUENCE 176 AA; 18933 MW; ECE369602A0ABD02 CRC64;
Query Match 3.4%; Score 6; DB 1; Length 176;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 6; Conservative 0; Mismatches 0; Indels 0;
QY 64 GSTVLK 59
Db 28 GSTVLK 33

```

Search completed: April 23, 2004, 14:07:23
Job time : 19 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

tein search, using sw model

April 23, 2004, 14:04:24 ; Search time 39 Seconds
(without alignments)
1448.148 Million cell updates/sec

US-09-147-801D-4

1 KDLMEIARMPVYKGERTEP.....QVWGNDAKFVPTTSNIW 179

OLIGO

Gapop 60.0 , Gapext 60.0

1017041 seqs, 315518202 residues

0

nits satisfying chosen parameters: 1017041

angth: 0

angth: 2000000000

Listing first 45 summaries

1: SPTEMBL 25: *
2: sp_archaea: *
3: sp_bacteria: *
4: sp_fungi: *
5: sp_human: *
6: sp_invertebrate: *
7: sp_mammal: *
8: sp_mhc: *
9: sp_organelle: *
10: sp_phage: *
11: sp_plant: *
12: sp_rodent: *
13: sp_virus: *
14: sp_vertebrate: *
15: sp_unclassified: *
16: sp_rvirus: *
17: sp_bacteriap: *
18: sp_archaeap: *

is the number of results predicted by chance to have a
per than or equal to the score of the result being printed,
ived by analysis of the total score distribution.

SUMMARIES

* Query Match	Length	DB	ID	Description
49.2	507	12	071023	071023 clethrionom
43.0	2256	12	Q8JV19	Q8JV19 ljunjan vir
15.1	2253	12	Q8JV21	Q8JV21 ljunjan vir
15.1	2253	12	Q8JV20	Q8JV20 ljunjan vir
8.4	2254	12	Q8ON16	Q8ON16 ljunjan vir
7.3	2177	12	Q8BES5	Q8BES5 human parec
4.5	736	10	Q8RUS3	Q8RUS3 oryza sativ
3.9	214	2	Q51078	Q51078 nocardia la
3.9	215	16	Q8X8V4	Q8X8V4 streptomyce
3.9	294	16	Q88DV2	Q88DV2 pseudomonas
3.9	304	6	Q8BE26	Q8BE26 macaca fasc
3.9	319	4	Q00477	Q00477 homo sapien
3.9	334	4	Q9NR44	Q9NR44 homo sapien
3.9	334	4	Q9BU81	Q9BU81 homo sapien
3.9	345	10	Q7XAU2	Q7XAU2 arabidopsis
3.9	350	4	Q99420	Q99420 homo sapien

17	7	3.9	357	4	015338	015338 hc
18	7	3.9	359	4	P78410	P78410 hc
19	7	3.9	361	5	Q9U0G4	Q9U0G4 pa
20	7	3.9	370	10	Q7X1H6	Q7X1H6 c
21	7	3.9	379	11	Q8BH91	Q8BH91 tr
22	7	3.9	389	10	Q84X13	Q84X13 a
23	7	3.9	391	16	Q891B7	Q891B7 c
24	7	3.9	394	16	P73510	P73510 s
25	7	3.9	408	10	Q9FWY1	Q9FWY1 a
26	7	3.9	419	16	Q8PGP7	Q8PGP7 x
27	7	3.9	419	16	Q8PCS6	Q8PCS6 x
28	7	3.9	430	10	Q94JC8	Q94JC8 c
29	7	3.9	440	17	Q9HKQ1	Q9HKQ1 t
30	7	3.9	460	5	Q9VS33	Q9VS33 dr
31	7	3.9	473	16	Q99TX5	Q99TX5 s
32	7	3.9	473	16	Q8NWD8	Q8NWD8 s
33	7	3.9	473	16	Q8CP42	Q8CP42 s
34	7	3.9	480	16	Q9RSL6	Q9RSL6 d
35	7	3.9	487	2	Q46060	Q46060 cc
36	7	3.9	495	4	Q9HCY1	Q9HCY1 hc
37	7	3.9	500	2	Q8GPV4	Q8GPV4 ps
38	7	3.9	505	16	Q7UG18	Q7UG18 r
39	7	3.9	513	4	O00481	O00481 hc
40	7	3.9	523	16	Q7VSN0	Q7VSN0 f
41	7	3.9	569	16	Q8R9L8	Q8R9L8 t
42	7	3.9	584	4	O00478	O00478 hc
43	7	3.9	608	17	Q9HU9	Q9HU9 h
44	7	3.9	660	5	Q18843	Q18843 ca
45	7	3.9	673	11	Q9QVT4	Q9QVT4 tr

ALIGNMENTS

RESULT 1

071023	PRELIMINARY;	PRT;	507 AA.
ID	071023		
AC	071023;		
DT	01-AUG-1998	(TREMBLrel. 07, Created)	
DT	01-AUG-1998	(TREMBLrel. 07, Last sequence update)	
DT	01-OCT-2003	(TREMBLrel. 25, Last annotation update)	
DE	Polyprotein (Fragment).		
OS	Clethrionomys glareolus picornavirus.		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Picornavi		
OX	NCBI_TaxID=75201;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Ljungan 145SL;		
RX	MEDLINE=99160759; PubMed=10049824;		
RA	Niklasson B., Kinnunen L., Hornfeldt B., Horling J., Benemar C.		
RA	Olof Hedlund K., Mätskova L., Hyypi T., Winberg G.;		
RT	"A new picornavirus isolated from bank voles (Clethrionomys		
RT	glareolus)."		
RL	Virology 255:86-93(1999).		
DR	EMBL; AF020541; AAC1265.1; --		
DR	InterPro; IPR008975; Viral_cap_coat.		
FT	NON TER	507	
SQ	SEQUENCE	507 AA;	56274 MW; B9FEB34607D3C7A1 CRC64;

Query Match 49.2%; Score 88; DB 12; Length 507;
Best Local Similarity 100.0%; Pred. No. 7.3e-89;
Matches 88; Conservative 0; Mismatches 0; Indels 0;

QY	1	KDLMEIARMPVYKGERTEPGTNGYQWSTHSPINWVFDGGTHLEDPNLL	
Db	318	KDLMEIARMPVYKGERTEPGTNGYQWSTHSPINWVFDGGTHLEDPNLL	
QY	61	YWRGSTVLKLTIVYASTFNKGLRMAFFP 88	
Db	378	YWRGSTVLKLTIVYASTFNKGLRMAFFP 405	

RESULT 2

PRELIMINARY; PRT; 2256 AA.
 (TrEMBLrel. 22, Created)
 (TrEMBLrel. 22, Last sequence update)
 (TrEMBLrel. 25, Last annotation update)
 us.
 RNA positive-strand viruses, no DNA stage; Picornaviridae;
 s.
 172314;

OM N.A.
 L;
 54258; PubMed=11955639;
 M., Johansson S.;
 ic analysis of Ljungan virus and A-2 plaque virus, new
 the Picornaviridae.";
 85:61-70(2002).

OM N.A.
 L;
 53720; PubMed=12163611;
 Niklasson B., Maizel J., Gorbalenya A.E., Lindberg A.M.;
 Analysis of Three Ljungan Virus Isolates Reveals a New,
 ot lineage of the Picornaviridae with a Cluster of Two
 A Proteins.";
 6:8920-8930(2002).

OM N.A.
 L;
 Niklasson B., Gorbalenya A., Lindberg A.M.;
 DEC-2000) to the EMBL/GenBank/DBJ databases.
 724; F:RNA helicase activity; IEA.
 968; F:RNA-directed RNA polymerase activity; IEA.
 350; P:transcription; IEA.
 079; P:virial genome replication; IEA.

PR009003; Cys Ser trypsin.
 PR006005; RNA helicase.
 PR007095; RNA_pol_DS_PS.
 PR001205; RNA_pol_P3D.
 PR007094; RNA_pol_PSVir.
 PR008975; Viral_cap_coat.
 80; RNA_dep_RNA_pol; 1.
 10; RNA helicase; 1.
 2256 AA; 253507 MW; 61FB9677D217A8B6 CRC64;

ilarity 43.0%; Score 77; DB 12; Length 2256;
 Conservativity 0; Mismatches 0; Indels 0; Gaps 0;
 VCDIGLNNTEMTIPYTWGNMPTGSGVIGWLRIIDVNLRTYNSSSPNAVCILQVK 162
 VCDIGLNNTEMTIPYTWGNMPTGSGVIGWLRIIDVNLRTYNSSSPNAVCILQVK 485

NDAKFMVPTTSNIW 179
 NDAKFMVPTTSNIW 502

PRELIMINARY; PRT; 2253 AA.
 (TrEMBLrel. 22, Created)
 (TrEMBLrel. 22, Last sequence update)
 (TrEMBLrel. 25, Last annotation update)

us.
 RNA positive-strand viruses, no DNA stage; Picornaviridae;
 s.
 172314;

RN
 RP
 SEQUENCE FROM N.A.
 RC
 STRAIN=87-012;
 RX
 MEDLINE=21954258; PubMed=11955639;
 RA
 Lindberg A.M., Johansson S.;
 RT
 "Phylogenetic analysis of Ljungan virus and A-2 plaque virus, r
 members of the Picornaviridae.";
 RL
 Virus Res. 85:61-70(2002).
 [2]
 RP
 SEQUENCE FROM N.A.
 RC
 STRAIN=87-012;
 RX
 MEDLINE=22153720; PubMed=12163611;
 RA
 Johansson S., Niklasson B., Maizel J., Gorbalenya A.E., Lindber
 "Molecular Analysis of Three Ljungan Virus Isolates Reveals a N
 Close-to-Root Lineage of the Picornaviridae with a Cluster of 1
 Unrelated 2A Proteins.";
 RL
 J. Virol. 76:8920-8930(2002).
 [3]

RN
 RP
 SEQUENCE FROM N.A.
 RC
 STRAIN=87-012;
 RA
 Johansson S., Niklasson B., Gorbalenya A., Lindberg A.M.;
 RL
 Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR
 EMBL; AF327920; AAM46079.1; -;
 DR
 GO; GO:0003724; F:RNA helicase activity; IEA.
 DR
 GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR
 GO; GO:0006350; P:transcription; IEA.
 DR
 GO; GO:0019079; P:virial genome replication; IEA.
 DR
 InterPro; IPR009003; Cys_Ser_trypsin.
 DR
 InterPro; IPR006005; RNA_helicase.
 DR
 InterPro; IPR007095; RNA_pol_DS_PS.
 DR
 InterPro; IPR001205; RNA_pol_P3D.
 DR
 InterPro; IPR007094; RNA_pol_PSVir.
 DR
 InterPro; IPR008975; Viral_cap_coat.
 DR
 Pfam; PF00680; RNA_dep_RNA_pol; 1.
 DR
 Pfam; PF00910; RNA_helicase; 1.
 SQ
 SEQUENCE 2253 AA; 253466 MW; 3C581A32096FCD41 CRC64;

Query Match 15.1%; Score 27; DB 12; Length 2253;
 Best Local Similarity 100.0%; Pred. No. 1.7e-20;
 Matches 27; Conservativity 0; Mismatches 0; Indels 0;

QY 105 VCDIGLNNTEMTIPYTWGNMPTRG 131
 Db 428 VCDIGLNNTEMTIPYTWGNMPTRG 454

RESULT 4
 Q8JV20
 ID Q8JV20 PRELIMINARY; PRT; 2253 AA.
 AC Q8JV20;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Polyprotein.
 OS Ljungan virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornavi
 Parechovirus.
 OK NCBI_TaxID=172314;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=174F;
 RX MEDLINE=21954258; PubMed=11955639;
 RA Lindberg A.M., Johansson S.;
 RT "Phylogenetic analysis of Ljungan virus and A-2 plaque virus, n
 members of the Picornaviridae.";
 RL Virus Res. 85:61-70(2002).
 [2]
 RP
 SEQUENCE FROM N.A.
 RC STRAIN=174F;
 RX MEDLINE=22153720; PubMed=12163611;
 RA Johansson S., Niklasson B., Maizel J., Gorbalenya A.E., Lindber
 "Molecular Analysis of Three Ljungan Virus Isolates Reveals a N
 Close-to-Root Lineage of the Picornaviridae with a Cluster of T

1 Proteins.;"

5:8920-8930(2002).

DM N.A.

1. Niklasson B., Gorbalenya A., Lindberg A.M.;
DEC-2000) to the EMBL/GenBank/DBJ databases.
921; AAM46080.1; --
724; F:RNA helicase activity; IEA.
968; F:RNA-directed RNA polymerase activity; IEA.
350; P:transcription; IEA.
079; P:viral genome replication; IEA.
R009003; Cys Ser tryptase.
R000605; RNA helicase.
R007095; RNA pol DS PS.
R001205; RNA pol P3D.
R007094; RNA pol_PSVir.
R008975; Viral_cap_coat.
30; RNA_dep_RNA_pol; 1.
10; RNA_helicase; 1.
2253 AA; 253484 MW; 1A6B9B07F325C793 CRC64;

15.1%; Score 27; DB 12; Length 2253;
ilarity 100.0%; Pred.No. 1.7e-20;
Conservative 0; Mismatches 0; Indels 0; Gaps 0;

IGLNTFTFTPTWGNWRPTRG 131
|||||
IGLNTFTFTPTWGNWRPTRG 454

PRELIMINARY; PRT; 2254 AA.

(TREMBlrel. 24, Created)
(TREMBlrel. 24, Last sequence update)
(TREMBlrel. 25, Last annotation update)

RNA positive-strand viruses, no DNA stage; Picornaviridae;

72314;

DM N.A.

1854; PubMed=12655084;
S., Niklasson B., Tesh R.B., Shafran D.R.,
Rosa A.P.A., Lindberg A.M.;
Characterization of M146, an American isolate of Ljungan
reveals the presence of a new LV genotype.";
1. 84:837-844(2003).
89; AAO83985.1; --
24; F:RNA helicase activity; IEA.
68; F:RNA-directed RNA polymerase activity; IEA.
50; P:transcription; IEA.
79; P:viral genome replication; IEA.
R009003; Cys Ser tryptase.
R000605; RNA_helicase.
R007095; RNA pol_DS_PS.
R001205; RNA pol_P3D.
R007094; RNA pol_PSVir.
R008975; Viral_cap_coat.
10; RNA_dep_RNA_pol; 1.
0; RNA_helicase; 1.
254 AA; 253072 MW; 17569BE279DE06BCD CRC64;

8.4%; Score 15; DB 12; Length 2254;
ilarity 100.0%; Pred.No. 3.6e-07;
Conservative 0; Mismatches 0; Indels 0; Gaps 0;

TNKGRLMAFFP 89

|||||

Db 396 ASTFNKGLRMAFFP 410

RESULT 6

Q8BES5
ID Q8BES5 PRELIMINARY; PRT; 2177 AA.
AC Q8BES5;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Polyprotein.
OS Human parechovirus 3.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornavi
OC Parechovirus.
OX NCBI_TaxID=195055;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A308/99;
RA Ito M., Yamashita T., Tsuzuki H., Sakae K., Takeda N.;
RT "Isolation and Identification of a Novel Human Parechovirus."
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB084913; BAC23086.1; --
DR GO; GO:0003724; F:RNA helicase activity; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR InterPro; IPR009003; Cys_Ser_tryptase.
DR InterPro; IPR007053; NC.
DR InterPro; IPR00605; RNA_helicase.
DR InterPro; IPR007095; RNA pol_DS_PS.
DR InterPro; IPR001205; RNA pol_P3D.
DR InterPro; IPR007094; RNA pol_PSVir.
DR InterPro; IPR008975; Viral_cap_coat.
DR Pfam; PF04970; NC; 1.
DR Pfam; PF0680; RNA_dep_RNA_pol; 1.
DR Pfam; PF00910; RNA_helicase; 1.
FT CHAIN 1 289 VP0.
FT CHAIN 290 545 VP3.
FT CHAIN 546 771 VP1.
FT CHAIN 772 920 2A.
FT CHAIN 921 1042 2B.
FT CHAIN 1043 1371 2C.
FT CHAIN 1372 1488 3A.
FT CHAIN 1489 1508 3B.
FT CHAIN 1509 1708 3C.
FT CHAIN 1709 2177 3D.
SQ SEQUENCE 2177 AA; 245821 MW; 7868G6EACE1D72D1 CRC64;

Query Match 7.3%; Score 13; DB 12; Length 2177;
Best Local Similarity 100.0%; Pred.No. 5.9e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0;

Oy 141 VLNRLTYNSSPN 153

|||||

506 VLNRLTYNSSPN 518

RESULT 7

Q8RUS3
ID Q8RUS3 PRELIMINARY; PRT; 736 AA.
AC Q8RUS3;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Similar to SET1.
GN P0425G02.10 OR P0468B07.30.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyt
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.

Nipponbare;
 Matsumoto T., Yamamoto K.;
 va nipponbare (GA3) genomic DNA, chromosome 1, PAC
 302, "
 FEB-2001) to the EMBL/GenBank/DBJ databases.

DM N.A.
 Nipponbare;
 Matsumoto T., Yamamoto K.;
 va nipponbare (GA3) genomic DNA, chromosome 1, PAC
 307, "
 FEB-2001) to the EMBL/GenBank/DBJ databases.

ITY: CONTAINS 1 SET DOMAIN.

247; BAB85235.1; -

260; BAB89651.1; -

RUS3; -

PR003105; G9a

PR003616; PostSET.

PR007728; Pre-SET.

PR001214; SET.

33; Pre-SET; 1.

56; SET; 1.

32; YDG SRA; 1.

308; PostSET; 1.

317; SET; 1.

166; SRA; 1.

50868; POST SET; 1.

50867; PRE SET; 1.

50280; SET; 1.

736 AA; 80094 MW; 8B65DC9173F83E57 CRC64;

ilarity 4.5%; Score 8; DB 10; Length 736;

Conservative 0; Mismatches 0; Indels 0;

LNRLT 146

|||||

LNRLT 41

PRELIMINARY; PRT; 214 AA.

(TREMBLrel. 01, Created)

(TREMBLrel. 01, Last sequence update)

(TREMBLrel. 25, Last annotation update)

1 DNA-binding protein.

ctamdurans.

ctinobacteria; Actinobacteridae; Actinomycetales;

lineae; Pseudonocardiaaceae; Anycolatopsis.

913;

DM N.A.

FEB-1993) to the EMBL/GenBank/DBJ databases.

; CAAT9799.1; -

; S40254.

703; F:5-amino-6-(5-phosphoribosylamino)uracil red. .; IEA.

577; F:DNA binding; IEA.

231; P:vitamin B2 biosynthesis; IEA.

PR002734; RibD_C.

72; RibD_C; 1.

214 AA; 23753 MW; 1172F003E7EDD9B5 CRC64;

ilarity 3.9%; Score 7; DB 2; Length 214;

Conservative 0; Mismatches 0; Indels 0;

3PGGT 23

|||||

3PGGT 56

RESULT 9

Q9X8Y4 PRELIMINARY; PRT; 215 AA.
 AC Q9X8Y4;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Putative DNA-binding protein.
 GN SC03702 OR SCH35.22C.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteridae; Actinobacteriales; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RX MEDLINE=21996410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylo
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyce
 coelicolor A3(2).";
 RL Nature 417:141-147(2002).
 DR EMBL; AL939117; CAB44405.1; -
 DR PIR; T36637; T36637.
 DR GO; GO:0008703; F:5-amino-6-(5-phosphoribosylamino)uracil red.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0009231; P:vitamin B2 biosynthesis; IEA.
 DR InterPro; IPR002734; RibD_C.
 DR Pfam; PF01872; RibD_C; 1.
 DR DNA-binding; Complete proteome.
 KW SEQUENCE 215 AA; 24059 MW; 25D546CB69E4913E CRC64;

Query Match 3.9%; Score 7; DB 16; Length 215;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 17 RTEPGGT 23

|||||

DB 51 RTEPGGT 57

RESULT 10

Q88DY2 PRELIMINARY; PRT; 294 AA.
 AC Q88DY2;
 DT 01-JUN-2003 (TREMBLrel. 24, Created)
 DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Lipoprotein, putative.
 GN PP4686.
 OS Pseudomonas putida (strain KT2440).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=160488;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=2423060; PubMed=12534463;
 RA Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,
 RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes
 RA Brinkac L., Beanon M., DeBoy R.T., Daugherty S., Kolonay J.,
 RA Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.
 RA Chris Lee P., Holtzapple E., Scanlan D., Tran K., Moazzzez A.,
 RA Utterback T., Rizzo M., Lee K., Kosack D., Moestl D., Wedler H.,
 RA Lauber J., Stjepandic D., Hoheisel J., Straetz M., Heim S.,

Eisen J., Timmis K.N., Duesterhoeft A., Tuemmler B.,
; snome sequence and comparative analysis of the
ly versatile *Pseudomonas putida* KT2440.";
robabil. 4:799-808 (2002).
791; AAN70259.1; -.
51; -.
PRO07314; DUF399.
PRO00437; Prok_lipoprot_S.
87; DUF399; 1.
00013; PROKAR_LIPOPROTEIN; 1.
steome.
294 AA; 32403 MW; 03EC7E0F42F7108 CRC64;
ilarity 3.9%; Score 7; DB 16; Length 294;
Conservative 0; Mismatches 0; Indels 0; Gaps 0;
RTEPG 21
|||||
RTEPG 272

PRELIMINARY; PRT; 304 AA.

(TREMBLrel. 17, Created)
(TREMBLrel. 17, Last sequence update)
(TREMBLrel. 25, Last annotation update)
l protein.
icularis (Crab eating macaque) (Cynomolgus monkey).
etazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
theria; Primates; Catarrhini; Cercopithecidae;
inae; Macaca.
3541;

WM N.A.
al cortex;
ida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
ugano S., Hashimoto K.;
f full-length cDNA clones from macaque brain cDNA

APR-2001) to the EMBL/GenBank/DBJ databases.

120; BAB41154.1; -.

PRO03599; IG.

PRO07110; IG-like.

109; IG; 1.

10835; IG LIKE; 2.

. protein.

104 AA; 33103 MW; 962F3123610AD8A8 CRC64;

ilarity 3.9%; Score 7; DB 6; Length 304;
Conservative 0; Mismatches 0; Indels 0; Gaps 0;

|||||

|||||

|||||

PRELIMINARY; PRT; 319 AA.

(TREMBLrel. 04, Created)
(TREMBLrel. 04, Last sequence update)
(TREMBLrel. 25, Last annotation update)
1.

; (Human).
etazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
theria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ruddy D.A., Kronmal G.S., Lee V.K., Mintier G.A., Quintana L.,
RA Domingo R. Jr., Meyer N.C., Basava A., McClelland E., Fullan A.
RA Mapa F.A., Moore T., Thomas W., Loeb D.B., Harmon C., Tsuchihas
RA Wolff R.K., Schatzman R.C., Feder J.N.;
RT "A 1.1 megabase transcript map of the human hereditary hemochr
RT locus.";
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U0546; AAB53424.1; -.
DR InterPro; IPR003599; IG.
DR InterPro; IPR007110; IG-like.
DR SMART; SM00409; IG; 1.
DR PROSITE; PS50835; IG LIKE; 1.
SQ SEQUENCE 319 AA; 34734 MW; 1697EF524F558749 CRC64;

Query Match 3.9%; Score 7; DB 4; Length 319;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0;

OY 41 DGGIHL 47

|||||

|||||

|||||

RESULT 13

Q9NR44

ID Q9NR44

AC Q9NR44; PRELIMINARY; PRT; 334 AA.

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE Butyrophilin, subfamily 3, member A2.

GN BTN3A2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostom

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

[1] _

RP SEQUENCE FROM N.A.

RX MEDLINE=21100910; PubMed=11170752;

RA Rhodes D.A., Stammers M., Malcherek G., Beck S., Trowsdale J.;

RT "The Cluster of BTN Genes in the Extended Major Histocompatibil

RT Complex.";

RL Genomics 71:351-362 (2001).

DR EMBL; AF257505; AAF76140.1; -.

DR InterPro; IPR003599; IG.

DR InterPro; IPR007110; IG-like.

DR SMART; SM00409; IG; 1.

DR PROSITE; PS50835; IG LIKE; 1.

SQ SEQUENCE 334 AA; 36399 MW; 95CFB887DB0DBAE4 CRC64;

Query Match 3.9%; Score 7; DB 4; Length 334;

Best Local Similarity 100.0%; Pred. No. 55;

Matches 7; Conservative 0; Mismatches 0; Indels 0;

OY 41 DGGIHL 47

|||||

|||||

|||||

RESULT 14

Q9BU81

ID Q9BU81

AC Q9BU81; PRELIMINARY; PRT; 334 AA.

DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE Similar to butyrophilin, subfamily 3, member A2 (DJ45P21.5).

GN BTN3A2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi

Rutheria; Primates; Catarrhini; Hominidae; Homo.
:9606;

ROM N.A.

(SEP-2001) to the EMBL/GenBank/DBJ databases.

ROM N.A.
1, and Placenta;

R.;

(DEC-2001) to the EMBL/GenBank/DBJ databases.

:832; AAH02832.1; -

:917; CAA17277.2; -

:214; AAH20214.1; -

:PR003599; Ig.

:PR007110; Ig-like.

:409; IG; 1.

:50835; IG LIKE; 1.

334 AA; 36427 MW; 940519D57F95EE4 CRC64;

ilarity 3.9%; Score 7; DB 4; Length 334;

Conservative 100.0%; Pred.No. 55;

Conservative 0; Mismatches 0; Indels 0; Gaps 0;

:G1HLE 47

|||||

:G1HLE 165

PRELIMINARY; PRT; 345 AA.

(TrEMBLrel. 25, Created)

(TrEMBLrel. 25, Last sequence update)

(TrEMBLrel. 25, Last annotation update)

ve nucleoside transporter ENT8 splice variant.

thaliana (Mouse-ear cress).

Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

ta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

; Brassicales; Brassicaceae; Arabidopsis.

3702;

OM N.A.

X., Wang D.;

JUL-2003) to the EMBL/GenBank/DBJ databases.

:059; AAQ16125.1; -

345 AA; 38341 MW; 34678B7A0646F2EE CRC64;

ilarity 3.9%; Score 7; DB 10; Length 345;

Conservative 100.0%; Pred.No. 56;

Conservative 0; Mismatches 0; Indels 0; Gaps 0;

VIGWL 137

|||||

VIGWL 342

: April 23, 2004, 14:08:16

cs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

ein search, using sw model

April 23, 2004, 13:57:18 ; Search time 58 Seconds
(without alignments)

872.000 Million cell updates/sec

JS-09-147-801D-4

390

1 KDLMEIAMPVYKGERTEP.....QVKMGNDKFMVPTTSNIW 179

3LOSUM62

3apop 10.0 , Gapext 0.5

1586107 seqs, 282547505 residues

its satisfying chosen parameters:

1586107

length: 0

length: 2000000000

Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

A: Geneseq 29Jan04:*

1: Geneseqp1980s:*

2: Geneseqp1990s:*

3: Geneseqp2000s:*

4: Geneseqp2001s:*

5: Geneseqp2002s:*

6: Geneseqp2003as:*

7: Geneseqp2003bs:*

8: Geneseqp2004s:*

s the number of results predicted by chance to have a
er than or equal to the score of the result being printed,
ved by analysis of the total score distribution.

SUMMARIES

Query	Match	Length	DB	ID	Description
00.0	179	2	AAW56274	AAW56274	Picornavi
10.9	3457	2	AAW84560	AAW84560	Polyprote
10.5	3457	2	AAW82504	AAW82504	Large pol
10.2	226	2	AAW27129	AAW27129	Equine rh
10.2	2247	2	AAW27126	AAW27126	Equine rh
9.5	202	2	AAW84563	AAW84563	Coat prot
9.5	3443	2	AAW84559	AAW84559	Polyproti
9.3	934	1	AAW20016	AAW20016	Sequence
9.0	227	1	AAW10035	AAW10035	Sequence
8.9	2206	2	AAW2210	AAW2210	True type
8.8	1147	5	ABB76724	ABB76724	Foot and
8.7	2179	1	AAW60243	AAW60243	Sequence
8.6	211	1	AAW10032	AAW10032	Sequence
8.5	1227	4	AAU02046	AAU02046	B. thurin
8.4	274	6	ABM70789	ABM70789	Staphyloc
8.4	1376	2	AAW97835	AAW97835	Kaposi's
8.4	1376	2	AAW93601	AAW93601	Kaposi's
8.4	246	1	AAW90613	AAW90613	Portion o
8.4	854	1	AAW50287	AAW50287	Hepatitis
8.4	1077	2	AAW95559	AAW95559	A partial
8.4	1091	2	AAW32426	AAW32426	Translate
8.4	2227	2	AAW05697	AAW05697	Attenuate
8.4	2227	2	AAW34074	AAW34074	Hepatitis
8.4	2227	3	AAW18609	AAW18609	Amino aci
8.4	2227	3	AAW18607	AAW18607	Amino aci

26	83	8.4	2227	3	AAW18608	AAW18608
27	83	8.4	2227	5	ABG31729	ABG31729
28	83	8.4	2227	5	ABG31727	ABG31727
29	83	8.4	2227	5	ABG31728	ABG31728
30	83	8.4	2227	6	ABU08640	ABU08640
31	83	8.4	2227	6	ABU08641	ABU08641
32	83	8.4	2227	6	ABU08639	ABU08639
33	82	8.3	993	1	AAW50231	AAW50231
34	82	8.3	2227	1	AAW60066	AAW60066
35	82	8.3	2227	5	AAW19899	AAW19899
36	82	8.3	2227	7	ABW00350	ABW00350
37	81.5	8.2	643	2	AAW16797	AAW16797
38	81.5	8.2	1186	2	AAW16796	AAW16796
39	81.5	8.2	1227	2	AAW44321	AAW44321
40	81.5	8.2	1227	4	AAW19950	AAW19950
41	80.5	8.1	2209	1	AAW20037	AAW20037
42	80	8.1	993	1	AAW50116	AAW50116
43	78.5	7.9	191	6	ABU42254	ABU42254
44	78.5	7.9	360	2	AAW30267	AAW30267
45	78.5	7.9	537	4	ABG25335	ABG25335

ALIGNMENTS

RESULT 1

AAW56274

ID AAW56274 standard; protein; 179 AA.

XX AC AAW56274;

XX DT 27-AUG-2003 (revised)

XX DT 30-JUL-1998 (first entry)

XX XX

XX DE Picornavirus Ljungan 145SL partial structural protein.

XX KW Vaccine; prophylactic treatment; myocarditis; multiple sclerosis

XX KW diabetes mellitus; DM; sudden infant death syndrome.

XX OS Picornaviridae.

XX PN WO9811133-A1.

XX PD 19-MAR-1998.

XX PF 09-SEP-1997; 97WO-SE001515.

XX PR 11-SEP-1996; 96SE-00003305.

XX PA (NIKL/) NIKLASSON B.

XX PI Niklasson B;

XX DR WPI; 1998-207327/18.

XX PT New picornavirus group causing mammalian disease - useful for d

XX PT vaccine and medicine development, for treating or preventing e;

XX XX myocarditis, multiple sclerosis, diabetes mellitus etc.

XX PS Claim 4; Page 29; 37pp; English.

XX CC The present invention provides for a new group of picornaviruse

XX CC picornaviruses were isolated from bank voles. The present parti

XX CC structural protein is encoded by the polprotein encoding cDNA o

XX CC Picornavirus Ljungan 145SL. The invention also claims that vacc

XX CC be prepared which include, as an immunising or neutralising comp

XX CC the picornavirus (optionally in attenuated or "killed" form), an

XX CC including a subunit of the virus or DNA corresponding to the vi

XX CC genomic RNA. Medicaments including one of these components as a

XX CC ingredient are claimed to be useful for prophylactic or therape

XX CC treatment of diseases caused by the picornavirus in mammals, esp

XX CC humans, e.g. myocarditis, multiple sclerosis, diabetes mellitus

XX CC sudden infant death syndrome. (Updated on 27-AUG-2003 to correc

SVASLTDI--GELSSVATG---SWSTTSATNLMELNIHPTSCAIQNL-ITQTP-LSV 949

depending on the exact location of cp2, the MCDV genome, can
78 kD of protein 5' of the capsid proteins (for which there
is no animal picornavirus protein). The DNA is used for
assistance to MCDV or viruses to which MCDV infection or
provides cross-resistance, including maize dwarf mosaic virus
by or all of the three coat protein genes from MCDV can be
vide protection for plants. (Updated on 25-MAR-2003 to correct

57 AA;

10.5%; Score 103.5; DB 2; Length 3457;
ilarity 24.7%; Pred. No. 0.22;
Conservative 28; Mismatches 54; Indels 49; Gaps 8;

HTHSPINWFDGGH-----LEDMPLNLFSSCYNYWGSTVLKLVYASTENK 79
-TTSATN-LMDLNHPTSCAIONGLITQPLSLAHAFARGLSKISIFGASLFTR 974
LAAAVFV-----AKRGTMSLDEISGYHNVCCLLNGQQTTFDLDPYSGQDSFYVR 1030
-----GNWMPRTGSGVIGLWIRIDVNLRTYNSSSPNAVNCILQVMGND 166
FDISAHGDNFM-----ITRLHLVILDKLVMSANSNINFSVTILPGSD 1077

andard; protein; 226 AA.

(revised)
(first entry)

ovirus 1 (ERHV1) polypeptide VP3 amino acid sequence.

ovirus 1; ERHV1; foot-and-mouth disease virus; vaccine; horse;
antigen; polypeptide; enzyme-linked immunosorbent assay;
protein; VP3.

itis A virus.

l.

96WO-AU000815.

95AU-00007201.

MELBOURNE.

Crabb BS, Feng L;

41692/31.

35181.

ence of equine rhinovirus 1 - and derived proteins or virus-
les, useful in vaccines and as diagnostic agents.

3e 34; 60pp; English.

sequence represents the amino acid sequence of the VP3
the polypeptide of equine rhinovirus 1 (ERHV1). The taxonomic
RhV1 is unclear, as physicochemical studies have shown that
acid density and base composition of ERHV1 differs from other
s. To this end, the nucleotide sequence encoding the
of ERHV1 was deduced. Analysis of this sequence suggests that
re closely related to foot-and-mouth disease virus. Individual
ins can be used to make vaccines to protect horses (and

possibly other animals) against ERHV1. Oligonucleotide primers
can be used for diagnosis of ERHV1 or related viruses, while an
the ERHV1 polypeptide can be used to detect ERHV1-specific anti-
the blood, particularly in enzyme-linked immunosorbent assay. I
differentiate between infected animals and those vaccinated with
CC vaccines (the infected animals will have antibodies reactive with
capsid proteins but vaccinated animals will not). The coding DN
CC for the present sequence represents an individual gene of the v
CC can be expressed in host systems to produce recombinant protein
CC like particles containing the individual ERHV1 proteins e.g. VP
CC also be used as vectors for delivering therapeutic or other use
CC agents, including vaccinating epitopes from other pathogens or
CC reproductive hormones. (Updated on 27-AUG-2003 to correct OS fi
XX Sequence 226 AA;

Query Match 10.2%; Score 100.5; DB 2; Length 226;
Best Local Similarity 27.9%; Pred. No. 0.011;
Matches 36; Conservative 20; Mismatches 58; Indels 15;

OY 56 SSCYNYWGRGTVLKLTVYASTFNKGRRLMAFEP--IMMQGTORKKHKCLFVWCD
DB 94 SFFAQYRGLSNFNFIPTGAATKAKFLVAFVPPHSAAPKTRDEAMACHAVWD
OY 114 FEMTIPY-TWGNWM-----RPTRGSGVIGLWIRIDVNLRTYNSSSPNAVY----C
DB 154 FSENVYPSPADFMNAVISAERTVNVSCWGLQVYALTALT---STDIAVNSKGRV
OY 164 GNDAKFMPV 172
DB 211 GPDFSRLHP 219

RESULT 5

AAW27126

ID AAW27126 standard; protein; 2247 AA.

XX AAW27126;

XX 27-AUG-2003 (revised)

DT 11-FEB-1998 (first entry)

XX Equine rhinovirus 1 (ERHV1) polypeptide amino acid sequence.

XX Equine rhinovirus 1; ERHV1; foot-and-mouth disease virus; vacci

KW diagnosis; antigen; polypeptide; enzyme-linked immunosorbent

XX recombinant protein.

XX Equine rhinitis A virus.

XX Key Location/Qualifiers

FT Peptide 1..207

FT /label= L

FT Cleavage-site 207..208

FT Peptide 208..287

FT /label= VP4

FT Cleavage-site 287..288

FT Peptide 288..517

FT /label= VP2

FT Cleavage-site 517..518

FT Peptide 518..743

FT /label= VP3

FT Cleavage-site 743..744

FT Peptide 744..991

FT /label= VP1

FT Cleavage-site 991..992

FT Peptide 992..1007

FT /label= 2A

FT Cleavage-site 1007..1008

FT Peptide 1008..1143

FT /label= 2B

FT Cleavage-site 1143..1144

FT Peptide 1144..1458

/label= 2C
te 1158. .1159
1459. .1552
/label= 3A
te 1552. .1553
1553. .1577
/label= 3B
te 1577. .1578
1578. .1782
/label= 3C
te 1782. .1783
1783. .2246
/label= 3D

1.
;
; 96WO-AU0000815.
; 95AU-00007201.
V MELBOURNE.
; Crabb BS, Feng L;
41692/31.
85178.

ence of equine rhinovirus 1 - and derived proteins or viruses, useful in vaccines and as diagnostic agents.
3e 30-32; 60pp; English.

sequence represents the amino acid sequence of the of equine rhinovirus 1 (ERhV1). The taxonomic status of ERhV1 as physicochemical studies have shown that the nucleic acid base comparison of ERhV1 differs from other rhinoviruses. To he nucleotide sequence encoding the polyprotein of ERhV1 was alysis of this sequence suggests that ERhV1 is more closely foot-and-mouth disease virus. Individual ERhV1 proteins can be a vaccines to protect horses (and possibly other animals) V1. Oligonucleotide primers and probes can be used for f ERhV1 or related viruses, while antigens of the ERhV1 can be used to detect ERhV1-specific antibodies in the blood, y in enzyme-linked immunosorbent assay. They can differentiate acted animals and those vaccinated with ERhV1 vaccines (the imals will have antibodies reactive with non-capsid proteins ted animals will not). Fragments of the DNA sequence represent genes of the virus and can be expressed in host systems to ombinant proteins. Virus like particles containing the ERhV1 proteins, can also be used as vectors for delivering or other useful agents, including vaccinating epitopes from gene or reproductive hormones. (Updated on 27-AUG-2003 to field.)

47 AA;

10.2%; Score 100.5; DB 2; Length 2247;
ilarity 27.9%; Pred.No. 0.26; Mismatches 58; Indels 15; Gaps 5;
Conservative 20; Mismatches 58; Indels 15; Gaps 5;
CYNWVGSTVLKLTYYASTFNKGRMAFFP--IMMQGTQKXKHKCLFMVCDIGLNN 113
FFAQYRGLNENFTTGAATKAKFLVAFVPPHSAAPKTRDEAMACIHAWDVLNSA 670
MTIPY-TWGNM-----RPTGSGVIGWLRIDVLNRLTYNSSSPNANV----CQLQVM 163
FNVPYSPADPMAYVAERTVNVSGMLQVYALTALT---STDIAVNSKGRVLVA 727
DAKEWVP 172
DFSURHP 736

RESULT 6
AAW84563
ID AAW84563 standard; protein; 202 AA.
XX
AC AAW84563;
XX
DT 01-APR-1999 (first entry)
XX
DE Coat protein CP3 of MCDV-Tn.
XX
KW Maize chlorotic dwarf virus strain Tennessee; MCDV-Tn; MCDV-T; M coat protein; transgenic plant; resistance; MCDV-Tn infection;
XX
OS Maize chlorotic dwarf virus.
XX
PN US5866780-A.
XX
PD 02-FEB-1999.
XX
PF 04-APR-1995; 95US-00416603.
XX
PR 04-APR-1995; 95US-00416603.
XX
PA (NOVS) NOVARTIS FINANCE CORP.
XX
PI Habera L, Law M, Reddick BB;
XX
DR WPI; 1999-142034/12.
DR N-PSDB; AAX03393.
XX
PT cDNA of maize chlorotic dwarf virus strain MCDV-Tn - and chimera for producing e.g. MCDV-Tn-resistant transgenic plants.
XX
PS Claim 6; Col 141-142; 73pp; English.
XX
CC The present sequence represents coat protein 3 (CP3) of the maize chlorotic dwarf virus strain Tennessee (MCDV-Tn) genome. The CP1 acid corresponds to nucleotides 3156-3761 of the entire nucleotide sequence of the MCDV-Tn positive strand RNA genome (AAX03389). CC distinct from known MCDV-T and MCDV-M1 isolates. The MCDV-Tn coat proteins may be expressed in transgenic plants to confer resistance to MCDV-Tn infection. The coat proteins can also be produced recombinantly in E. coli.
XX
SQ Sequence 202 AA;
Query Match 9.5%; Score 94.5; DB 2; Length 202;
Best Local Similarity 22.4%; Pred. NO. 0.045;
Matches 38; Conservative 32; Mismatches 59; Indels 41;
QY 29 WSHT-----HSPINWVFDGIIHLEDMFNLNLFSSCYNWVGSTVLKLT
DB 21 WSTSSLLQLQNVHPTACFISDG---LVTOTPLSVIAHAFARMKGSXKFTTIF
QY 79 KGRLEMAFFPIMMQGTQK-----HKCLFMVCDIGLNNFTFMTIPY-T
DB 78 RGRVLVAIPV-----AKKETLTIEISGYHNVNCLL--NGERTSFELEVPHS
QY 128 PTRGS-----VIGWLRIDVLNRLTYNSSSPNANVILQVMGND 166
DB 132 VCRDALFDVSSYAQNFMITRLHMVVIDTLVNSSNASNTISYCVMMGPCKD 181
RESULT 7
AAW84559
ID AAW84559 standard; protein; 3443 AA.
XX
AC AAW84559;
XX
DT 31-MAR-1999 (first entry)
XX
DE Polyprotein encoded by the MCDV-Tn positive strand RNA genome.


```

cyclic dwarf virus strain Tennessee; MCDV-Tn; MCDV-T; MCDV-M1;
1; transgenic plant; resistance; MCDV-Tn infection.
cyclic dwarf virus.
:
:
: 95US-00416603.
: 95US-00416603.
ARTIS FINANCE CORP.
Law M, Reddick BB;
12034/12.
13389.
cyclic chlorotic dwarf virus strain MCDV-Tn - and chimeric genes
ing e.g. MCDV-Tn-resistant transgenic plants.
Col 75-92; 73pp; English.
sequence represents the polyprotein encoded by the maize
cyclic dwarf virus strain Tennessee (MCDV-Tn) genome. MCDV-Tn is
from known MCDV-T and MCDV-M1 isolates. The specification also
includes sequences for 3 MCDV-Tn coat proteins (see AAW84561-63),
which are expressed in transgenic plants to confer resistance to MCDV-
1. The coat proteins can also be produced recombinantly
43 AA;
9.5%; Score 94.5; DB 2; Length 3443;
Similarity 22.4%; Pred. No. 2.2;
Conservative 32; Mismatches 59; Indels 41; Gaps 7
TT-----HSPINWVFDGGIHDMPNLNLFSCYNYMKGSTVLKLTIVYASTFN 78
TTSLLMLQLNWHPHTACFISDG--LVTQTPLSVIAHAFARWGRGSLKFTITFGASMT 984
RLRMAFFPIMMGTQTKK-----HKCLPMWCDIGLNTTFEMTIPY-TWGNWNR 127
: : : : : : : : : : : : : : : : : : : : : : : : : :
RVLVAALP-----AKREKLTITTEISGYNVNMCCL--NGERTSFELEVPYHVSVDGDSY 1038
RGS-----VIGWLRIDVNLRLTYNSSSPNAVNCILQVKGND 166
: : : : : : : : : : : : : : : : : : : : : : : : : :
DALFDVSSYAQNFMITRLHMVVVIDTLVMSSNASNTISYCVMMGPCKD 1088
indard; protein; 934 AA.
(revised)
(revised)
(first entry)
p20, VP4, VP2, VP3, VP1 and p52 encoded by recombinant
AS1/t 76.
body; capsid protein; immunogen; antigen;
with disease; ss.
with disease virus.
Location/Qualifiers
1..75
/label= p20

```

(revised)
(revised)
(first entry)

VP3 encoded by part of the sequence of DNA insert FMDV-1448.

ot and mouth disease; antigen.

uth disease virus.

Location/Qualifiers

9..227

/note="encoded by AAN10026"

; 80GB-00015635.

; 80GB-00015635.

; 80GB-00015655.

; 80GB-00026661.

; 80GB-00028983.

; 81GB-00014309.

GEN NV.

GEN NV.

PH, Kupper HA, Schaller H, Keller W;

1439D/50.

10026.

(s) with foot and mouth disease antigen specificity - produced
quences by transformed hosts etc.

g 11; 90pp; English.

rs claim DNA sequences that encode antigenic polypeptides of
ed from FMDV-115, FMDV-144, FMDV-1034, FMDV-1448, FMDV-1824,
VP1-1, VP1-5 FMDV-1034-Bal or FMDV-1034-Bal(EcoRI-HindIII). In
FMDV serotypes O, A, C, SAT 1, SAT 2, SAT 3, and Asian type I.
nic polypeptides are also claimed. (Updated on 25-MAR-2003 to
field.) (Updated on 25-MAR-2003 to correct PA field.) (Updated
003 to standardise OS field)

7 AA;

ilarity 9.0%; Score 89.5; DB 1; Length 227;

Conservative 27; Mismatches 67; Indels 21; Gaps 7;

GGIHEMPNINL--FSSCYNYWGSTVLKLTIVYASTFNKGRLEMAFFPIMQ--GTQ 95

MSLAQKQMSNTFLAGLAQYTYQSGTINLHFMXTGPTDAKARYWVAVAPLGMPEPKTP 136

XHKCLFWVCDIGLNNTEMTIPY-TWGNWMPTEG-----SVIGWLRIDVNLRLTYNS 149

AAHCIIHAEWDGLNSKFTFSIPYLSAADYATASGVAETTNVQGV---CLFQITHGK 193

PNAVNCIIQVWNGNDKFWVP-----TTS 175

GDAL--VVLASAGKDFELRLPVDARAETTS 223

andard; protein; 2206 AA.

XX 25-MAR-2003 (revised)
DT 14-JUL-1992 (first entry)
XX True type 3 poliovirus protein from LED3.
XX RNA virus; error reduction.
XX Poliovirus.
XX OS
XX WO9203538-A.
XX 05-MAR-1992.
XX 20-AUG-1991; 91WO-US005890.
XX 20-AUG-1990; 90US-00569916.
XX 20-AUG-1990; 90US-00570000.
XX (UYCO) UNIV COLUMBIA NEW YORK.
XX Racaniello V, Ta, Tatem JM, Weekslevy CL;
XX WPI; 1992-096882/12.
XX DR N-PSDB; AAQ22965.
XX New vaccine against infectious polio-virus comprises RNA virus
PT producing RNA virus cDNA and viable RNA virus.
XX Disclosure; Fig 6; 110pp; English.
XX The protein sequence was deduced from the cDNA sequence of P3 f
CC obtd as in AAQ22965. The cDNA sequence is that of a true RNA vi
CC the cDNA directs the prodn. of a viable RNA virus which is phen
CC similar to the source virus. The full length cDNA in pLED3 was
CC infectious. In vitro transcription of pLED3 cDNA using T7 RNA p
CC produced RNAs which possessed several erroneous amino acids. Th
CC viruses are used in vaccines against polio. The screening metho
CC used during amplification of the source virus for vaccine prodn
CC ensure maintenance of C at position 2493 in the viral genome i.
CC increasing the attenuation. The new prod. overcomes the problem
CC introduced during replication of ss RNA, which is much higher t
CC DNA. (Updated on 25-MAR-2003 to correct PA field.)
XX SQ Sequence 2206 AA;
Query Match 8.9%; Score 88.5; DB 2; Length 2206;
Best Local Similarity 23.0%; Pred. No. 5.7; Indels 5;
Matches 26; Conservative 25; Mismatches 57;
QY 59 YNYWRGSLVTLKLTIVYASTFNKGRLEMAFFPIMQGTORKKKKCL--FMVCDIGL
Db 447 YTHWAGSLKFTFLFCGSMWATGKILVAVAPGAQPTSRKEAMLGTHVICDLGL
QY 117 TTPYTWG-NWMPTEGSGVI--GWLRLDVLNRLTYNSSSPNAVNCILQVWNGND
Db 507 VVPVSNVTYRTQDSFTEGGYISMFYQTRIIVPLSTPKMSMLGFSVACND
RESULT 11
ABB76724
ID ABB76724 standard; protein; 1147 AA.
XX ABB76724;
XX ABB76724;
DT 29-AUG-2003 (revised)
DT 29-MAY-2002 (first entry)
XX Foot and Mouth Disease Virus Pl, 2A and 3C protein sequence.
XX Virucide; vaccine; foot and mouth disease; Pl region; capsid;
XX 3C protease.

ith disease virus.

Location/Qualifiers
2..737
/label= P1
738..753
/label= 2A
913..1126
/label= 3C

-AL.

; 2001WO-FR002042.

; 2000FR-00008437.

IAL.

rman A, Audonnet J, Lombard M;

30837/17.

53056.

ent effective vaccines against foot-and-mouth disease,
eombinantly produced empty virus capsids as antigens.

Fig 3; 79pp; French.

invention relates to a vaccine against foot and mouth disease
ising (in addition to a veterinary vehicle or excipient) an
sisting of empty FMD virus capsids, obtained by expression in
cells of the CDNA of the following regions of the FMD genome:
on encoding the capsid and the region encoding the 3C
ne vaccine is effective, reliable and stable, and is effective
s. The vaccine is useful against foot and mouth disease,
in cows, sheep, pigs or goats. The present sequence is a
cm FMD comprising the P1, 2A and 3C regions, which was used to
the invention. (Updated on 29-AUG-2003 to standardise OS

47 AA;

ilarity 22.6%; Pred.No. 3.4; Length 1147;

Conservative 29; Mismatches 64; Indels 20; Gaps 6;

GGIHLEDPN--LNLFSQYNYMRGSGTVLKLTVYASTFNKGRRLMAFFPIMMQ---GT 94

VSLAAKHSNTYLSGIAQYTYQSGTINLHFMFTGSTDASKARYVMVAIPPQVETPPDT 440

KKHKLFWCDIGLNNFTFWTIP-----YTWGNMWRPTRGSGVIGLWIDVLNRLT 146

EAACHIAEWDITGLNSKFTFSIPYSAADYATASDTAETT--NVQGWV---CVYQIT 495

SSSPNAVNCILQVMGNDKFMVP 172

KAENDTL--LVSASAGKDFELRLP 519

andard; protein; 2179 AA.

(revised)
(first entry)

coding the entire genomic RNA of human rhinovirus.

antibody; MAB; HRV; vaccine; ss.

XX OS Human rhinovirus sp.
XX FH Key Location/Qualifiers
XX FT Region 1..69
XX FT Region /label= VP4 structural protein
XX FT Region 70..331
XX FT Region /label= VP2 structural protein
XX FT Region 332..567
XX FT Region /label= VP3 structural protein
XX FT Region 568..856
XX FT Region /label= VP1 structural protein
XX FT Region 857..1002
XX FT Region /label= 3B protein
XX FT Region 1003..1099
XX FT Region /label= 5B protein
XX FT Region 1100..1429
XX FT Region /label= X protein
XX FT Region 1430..1514
XX FT Region /label= protein 1B
XX FT Region 1515..1537
XX FT Region /label= protein VPg
XX FT Region 1538..1719
XX FT Region /label= Protease
XX FT Region 1720..2179
XX FT Region /label= Replicase
XX EP169146-A.
XX PD 22-JAN-1986.
XX PF 17-JUL-1985; 85EP-00401465.
XX PR 20-JUL-1984; 84US-00632785.
XX PR 10-APR-1985; 85US-00721735.
XX PA (MERI) MERCK & CO INC.
XX PI Colonne RJ, Mitzutani S;
XX WPI; 1986-022809/04.
XX DR N-PSDB; AAN60194.
XX PT New DNA encoding the entire genomic RNA of human rhinovirus 14
XX PT monoclonal antibodies which block attachment or neutralise infe
XX PS rhinovirus.
XX Example 11; Page 22-39; 80pp; English.
XX Sequence may be used for the manufacture of hybridoma cells exp
CC the HRV or fragments thereof. The fusion products may be used i
CC immunisation, or to raise MABs for passive treatment of HRV inf
CC (Updated on 24-OCT-2003 to standardise OS field)
XX SQ Sequence 2179 AA;
Query Match 8.7%; Score 86; DB 1; Length 2179;
Best Local Similarity 31.5%; Pred. No. 11;
Matches 23; Conservative 11; Mismatches 37; Indels 2;
QY 59 YNYWRGSTVLKLTIVYASTFNKGRRLMAFFPIMMQTQRKKHKL--FMVCDIGL
Db 435 YTHWSGSLRFSMYTGPALSSAKLTAYTPPGARGPQDRREAMLGHVVDIGL
QY 117 TIPYTWGNMWRPT 129
Db 495 TIATSGVQFRYT 507
RESULT 13
AAP10032
ID AAP10032 standard; protein; 211 AA.
XX

(first entry)
us aureus protein #29.
d; vaccine; gene therapy; infection; sepsis; diagnosis;
say; antibiotic target.
us aureus.
A2.
: 2002WO-IB002637.
: 2001GB-00007661.
RON SPA.
Mora M, Scarselli M;
0786/11.
72349.
ococcus aureus protein, useful as a vaccine for treating or
Staphylococcal infection, specifically an infection caused by
S. aureus.
ID NO 58; 49pp; English.
on relates to novel genes and encoded proteins from
us aureus. A composition comprising the S. aureus protein, a
i encoding the protein, or an antibody to the protein, is
pharmaceutical, particularly as a vaccine for treating or
infection due to Staphylococcus bacteria, specifically an
used by S. aureus. The composition is particularly useful for
preventing sepsis in a patient. The composition can also be
agnostics. The protein is also used in an assay for enzymatic
as a target for antibiotics. This sequence represents one of
S. aureus proteins of the invention
AA;
8.4%; Score 83.5; DB 6; Length 274;
ilarity 25.8%; Pred. No. 1.2;
Conservative 16; Mismatches 42; Indels 31; Gaps 5;
ARMSVYKGRTEPGGTNGYFOWSHTHSPINWVFDGGIHLEDPNINLFSSCYNWYRG 64
EVSINLYKGYSESVMWIDSEFLDS-----NMWFENGIPLDFTPKYTHTSNQFTIWN 174
JLKLTIVASTFNKGRLEMAFFPIMMOGTQRKKHKLFWVCDIGLNNTFTFTIYTWGN 124
-----DTIN-----PRFKHD-LKILINLNASGGFEL-VNYTTGD 208
: April 23, 2004, 13:58:32
CS

Query #	Match	Length	DB	ID	Description
10-9	3457	2	US-08-416-603-4	Sequence 4, Appl	
10-9	226	3	US-09-091-219-5	Sequence 5, Appl	
10-2	226	4	US-09-660-541-5	Sequence 5, Appl	
10-2	2232	3	US-09-091-219-25	Sequence 25, Appl	
10-2	2232	4	US-09-660-541-25	Sequence 25, Appl	
10-2	2247	3	US-09-091-219-2	Sequence 2, Appl	
10-2	2247	4	US-09-660-541-2	Sequence 2, Appl	
9-5	202	2	US-08-416-603-10	Sequence 10, Appl	
9-5	3443	2	US-09-416-603-2	Sequence 2, Appl	
8-5	1227	4	US-09-661-322A-63	Sequence 63, Appl	
8-5	2318	3	US-09-091-219-4	Sequence 24, Appl	
8-5	2318	4	US-09-660-541-24	Sequence 24, Appl	
8-4	1376	1	US-08-420-235B-3	Sequence 3, Appl	
8-4	1376	3	US-08-793-624-3	Sequence 3, Appl	
8-4	1376	5	PCT-US95-10194-3	Sequence 3, Appl	
8-4	1091	6	551630-2	Patent No. 551630	
8-4	2227	3	US-08-475-886-2	Sequence 2, Appl	
8-4	2227	3	US-08-475-886-4	Sequence 4, Appl	
8-4	2227	3	US-08-475-886-6	Sequence 6, Appl	
8-4	2227	3	US-08-397-232-2	Sequence 2, Appl	
8-4	2227	3	US-08-397-232-4	Sequence 4, Appl	
8-4	2227	3	US-09-171-387-2	Sequence 2, Appl	
8-4	2227	4	US-09-653-439-2	Sequence 2, Appl	
8-4	2227	4	US-09-653-439-4	Sequence 4, Appl	
8-4	2227	4	US-09-653-439-6	Sequence 6, Appl	
8-4	2227	4	US-10-135-988-2	Sequence 2, Appl	
8-4	2227	4	US-10-135-988-4	Sequence 4, Appl	

09:21:17 2004

us-09-147-801d-4.rai

HAFARWRSGLKISIIFGASLFRGRILLAAVAVP-----AKRKGTMSLDELISGYHNVCCL 1005
GLNTEEMTIDYTW-----CNWWRPTRGSGVIGWLRIDVLRNLT 147
GQOTTFEIEIPYVSGQDSFYRDALFDSAHGDNFM-----ITRLHLVILDKVM 1058
SSPNVAVNCILOVKMGND 166
NASNSINFSVTLGPGSD 1077
plication US/09091219
1592
ATION:
UDDERT, Michael J.
ABB, Brendan S.
NG, Li
NTION: EQUINE RHINOVIRUS 1 PROTEINS
E: 040268/0151
CATION NUMBER: US/09/091,219
G DATE: 1998-10-05
CATION NUMBER: PCT/AU96/00815
G DATE: 1996-12-18
CATION NUMBER: AU PN7201
G DATE: 1995-12-18
ID NOS: 25
entIn Ver. 2.0

uine rhinovirus 1

10.2%; Score 100.5; DB 3; Length 226;
ilarity 27.9%; Pred. No. 0.00089;
Conservative 20; Mismatches 58; Indels 15; Gaps 5;
CYNWVRGSLVTLKLVYASTENKGRLMARFP--IMMQGTQKXKHKCLFMVCDIGLNT 113
FFAQYRSLNFNFTGAAATKAKFLVAFVPPHSAAPKTRDEAMACHVAVMDVGLNSA 153
MTIPY-TWGNM-----RPTRGSGVIGWLRIDVLRNLTYSNPNVAVN---CILQVKM 163
FNVPYSPADFMAYSAERTVNVVSGWLVYALTALT---STDIANSKGRVAVNSA 210
DAKFWP 172
DPSLRHP 219

plication US/09660541

1136
ATION:
UDDERT, Michael J.
RAB, Brendan S.
ENG, Li
NTION: EQUINE RHINOVIRUS 1 PROTEINS
E: 040268/0151
CATION NUMBER: US/09/660,541
G DATE: 2000-09-12
TION NUMBER: EARLIER APPLICATION NUMBER: 09/091,219
DATE: EARLIER FILING DATE: 1998-10-05
TION NUMBER: EARLIER APPLICATION NUMBER: AU PN7201
DATE: EARLIER FILING DATE: 1995-12-18
ID NOS: 25
entIn Ver. 2.0

ORGANISM: equine rhinovirus 1
US-09-660-541-5
Query Match 10.2%; Score 100.5; DB 4; Length 226;
Best Local Similarity 27.9%; Pred. No. 0.00089;
Matches 36; Conservative 20; Mismatches 58; Indels 15;
QY 56 SSCYNWVRGSLVTLKLVYASTENKGRLMARFP--IMMQGTQKXKHKCLFMVCD
DB 94 SSFFAQYRSLNFNFTGAAATKAKFLVAFVPPHSAAPKTRDEAMACHVAVMD
QY 114 FEMTIPY-TWGNM-----RPTRGSGVIGWLRIDVLRNLTYSNPNVAVN---C
DB 154 FSNVPYSPADFMAYSAERTVNVVSGWLVYALTALT---STDIANSKGRV
QY 164 GNDKFWP 172
DB 211 GPDFSRLHP 219

RESULT 4

US-09-091-219-25
; Sequence 25, Application US/09091219
; Patent No. 6171592
; GENERAL INFORMATION:
; APPLICANT: STUDDERT, Michael J.
; APPLICANT: CRABB, Brendan S.
; APPLICANT: FENG, Li
; TITLE OF INVENTION: EQUINE RHINOVIRUS 1 PROTEINS
; FILE REFERENCE: 040268/0151
; CURRENT APPLICATION NUMBER: US/09/091,219
; CURRENT FILING DATE: 1998-10-05
; EARLIER APPLICATION NUMBER: PCT/AU96/00815
; EARLIER FILING DATE: 1996-12-18
; EARLIER APPLICATION NUMBER: AU PN7201
; EARLIER FILING DATE: 1995-12-18
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 2232
; TYPE: PRT
; ORGANISM: equine rhinovirus 1
US-09-091-219-25

Query Match 10.2%; Score 100.5; DB 3; Length 2232;
Best Local Similarity 27.9%; Pred. No. 0.024; 58; Indels 15;
Matches 36; Conservative 20; Mismatches 58; Indels 15;
QY 56 SSCYNWVRGSLVTLKLVYASTENKGRLMARFP--IMMQGTQKXKHKCLFMVCD
DB 611 SSFFAQYRSLNFNFTGAAATKAKFLVAFVPPHSAAPKTRDEAMACHVAVMD
QY 114 FEMTIPY-TWGNM-----RPTRGSGVIGWLRIDVLRNLTYSNPNVAVN---C
DB 671 FSNVPYSPADFMAYSAERTVNVVSGWLVYALTALT---STDIANSKGRV
QY 164 GNDKFWP 172
DB 728 GPDFSRLHP 736

RESULT 5

US-09-660-541-25
; Sequence 25, Application US/09660541
; Patent No. 6531136
; GENERAL INFORMATION:
; APPLICANT: STUDDERT, Michael J.
; APPLICANT: CRABB, Brendan S.
; APPLICANT: FENG, Li
; TITLE OF INVENTION: EQUINE RHINOVIRUS 1 PROTEINS
; FILE REFERENCE: 040268/0151
; CURRENT APPLICATION NUMBER: US/09/660,541
; CURRENT FILING DATE: 2000-09-12

; Patent No. 6531136
 ; GENERAL INFORMATION:
 ; APPLICANT: STUDDERT, Michael J.
 ; APPLICANT: CRABB, Brendan S.
 ; APPLICANT: FENG, Li
 ; TITLE OF INVENTION: EQUINE RHINOVIRUS 1 PROTEINS
 ; FILE REFERENCE: 040268/0151
 ; CURRENT APPLICATION NUMBER: US/09/660,541
 ; CURRENT FILING DATE: 2000-09-12
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/091,219
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1998-10-05
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: AU PN7201
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1995-12-18
 ; NUMBER OF SEQ ID NOS: 25
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 2
 ; LENGTH: 2247
 ; TYPE: PRT
 ; ORGANISM: equine rhinovirus 1
 ; US-09-660-541-2

10.2%; Score 100.5; DB 4; Length 2232;
 ilarity 27.9%; Pred. No. 0.024;
 Conservative 20; Mismatches 58; Indels 15; Gaps 5;

Qy	56	SSCYNVW	RGSTVL	KLVTV	ASTNKG	RLRMAPP	--IM	OGTQRK	KKHCL	FMVCD
Db	611	SSFPQY	RGSLNF	IFTGA	AAATK	AKELV	AFVPPH	SAAPK	TEDEAM	ACHAVWD
Qy	114	FEMTTP	YTWGNW	-----	RPTRG	SVIGML	RIDL	VNLRL	TYNSSP	NAVN----
Db	671	FSFNVP	YPSPAD	EMAVYS	AERTV	VNVV	SGML	QVYALT	ALT---	STDI
Qy	164	GNDAK	FMVP	172						
Db	728	GPDFS	LRHP	736						

ADDERT, Michael J.
ABB, Brendan S.
VG, Li
ATTION: EQUINE RHINOVIRUS 1 PROTEINS

```

RESULT 8
US-08-416-603-10
; Sequence 10, Application US/08416603
; Patent No. 5866780
; GENERAL INFORMATION:
; APPLICANT: Law, Marcus
; APPLICANT: Hebbara, Ledare
; APPLICANT: Reddick, Bradford B.
; TITLE OF INVENTION: Maize Chlorophyll
; TITLE OF INVENTION: Uses Therefor
;

```

clarity	10.2%;	Score 100.5;	DB 3;	Length 2247;
Conservative	27.9%;	Pred. No. 0.024;		
		20; Mismatches	58; Indels	15; Gaps
				5;

```

/
/ COMPUTER READABLE FORM:
/
/ MEDIUM TYPE: Floppy disk
/
/ COMPUTER: IBM PC compatible
/
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/
/ SOFTWARE: Patentin Release #1.0, Version #1.30B
/
/ CURRENT APPLICATION DATA:
/
/ APPLICATION NUMBER: US/08/416,603
/

```

ATTORNEY/AGENT INFORMATION:
NAME: Lloyd, Jeffrey
REGISTRATION NUMBER: 35,589
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 202 amino acids

Publication US/09660541

RVLVAAIPV-----AKRKETLTIEEISGYHNMCLL--NGERTSFELEVPY

Best Local Similarity 24.3%; Pred

Query Match	8.5%;	Score 84.5;	DB 3;	Length 2318;
Best Local Similarity	24.3%;	Pred. No. 2;		
Matches 37;	Conservative 26;	Mismatches 68;	Indels 21;	

us-09-147-801d-4.rai

[illegible]

plication PC/TUS9510194
MATION:
The trustees of Columbia University in the City of New York
City
VENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
VENTION: SEQUENCES AND USES THEREOF
SEQUENCES: 45
NCE ADDRESS:
: Cooper & Dunham LLP
1185 Avenue of the Americas
w York
ew York
U.S.A.
36
ADABLE FORM:
PE: Floppy disk
IBM PC compatible
SYSTEM: PC-DOS/MS-DOS
PatentIn Release #1.25
LOCATION DATA:
ON NUMBER: PCT/US95/10194

TE: 1
ACTION: 1
AGENT INFORMATION: 1
ite, John P. 1
ION NUMBER: 28, 678 1
/DOCKET NUMBER: 45185-C-PCT/JPW/MSC 1
ACTION INFORMATION: 1
: (212) 278-0400 1
(212) 391-0525 1
OR SEQ ID NO: 3: 1
ABACTERISTICS: 1
1376 amino acids 1
ino acid 1
linear 1
PE: protein 1

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

tein search, using sw model

April 23, 2004, 13:59:44 ; Search time 42 Seconds
(without alignments)
1178.311 Million cell updates/sec

US-09-147-801B-4

990
1 KDLMEIARMPVYKGERTEP.....QVKGNDGAKFMVPTTSNIW 179

BLOSUM62

Gapop 10.0 , Gapext 0.5

1133595 seqs, 276475211 residues

hits satisfying chosen parameters: 1133595

length: 0

ength: 2000000000

Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Published Applications AA.*

- 1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/prodata/2/pubpaa/PCR_NEW_PUB.pep.*
- 3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/prodata/2/pubpaa/PCR_NEW_PUBCOMB.pep.*
- 7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/prodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/prodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/prodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/prodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/prodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pep.*

is the number of results predicted by chance to have a
ter than or equal to the score of the result being printed,
ived by analysis of the total score distribution.

SUMMARIES

Query Match	Length	DB	ID	Description
100.0	179	10	US-09-147-801B-4	Sequence 4, Appli
41.8	2179	14	US-10-224-999A-3481	Sequence 3481, Ap
8.8	1147	15	US-10-327-481A-38	Sequence 38, Appl
8.5	1227	15	US-10-428-961-63	Sequence 63, Appl
8.5	1108	15	US-10-369-493-22044	Sequence 22044, A
8.4	2227	13	US-10-135-988-2	Sequence 2, Appli
8.4	2227	13	US-10-135-988-4	Sequence 4, Appli
8.4	2227	13	US-10-135-988-6	Sequence 6, Appli
8.3	248	14	US-10-272-459-43	Sequence 43, Appl
8.3	836	14	US-10-272-459-40	Sequence 40, Appl
8.3	980	14	US-10-272-459-41	Sequence 41, Appl
8.3	2227	9	US-09-929-955-12	Sequence 12, Appl
8.3	2227	13	US-10-104-966-12	Sequence 12, Appl
8.2	643	9	US-09-826-660-25	Sequence 25, Appl
8.2	1186	9	US-09-826-660-23	Sequence 23, Appl

16	79	8.0	4852	12	US-10-412-406-33	Sequenc
17	78.5	7.9	191	12	US-10-282-122A-70178	Sequenc
18	78.5	7.9	232	14	US-10-237-386-64	Sequenc
19	75	7.6	457	14	US-10-225-567A-469	Sequenc
20	75	7.6	457	15	US-10-292-798-618	Sequenc
21	74.5	7.5	754	15	US-10-369-493-5374	Sequenc
22	74	7.5	495	15	US-10-295-027-875	Sequenc
23	74	7.5	631	9	US-09-789-919-56	Sequenc
24	74	7.5	683	9	US-09-738-626-6961	Sequenc
25	74	7.5	683	12	US-10-450-055-26	Sequenc
26	74	7.5	720	13	US-10-047-412A-8	Sequenc
27	73.5	7.4	161	12	US-10-282-122A-68575	Sequenc
28	72.5	7.3	146	12	US-10-424-599-211434	Sequenc
29	72.5	7.3	742	10	US-09-374-046A-160	Sequenc
30	72.5	7.3	742	12	US-10-616-263-160	Sequenc
31	72	7.3	226	14	US-10-237-386-63	Sequenc
32	72	7.3	701	14	US-10-156-761-9875	Sequenc
33	72	7.3	756	15	US-10-369-493-11983	Sequenc
34	72	7.3	1953	15	US-10-369-493-1945	Sequenc
35	71.5	7.2	507	12	US-10-282-122A-72657	Sequenc
36	71	7.2	257	9	US-09-893-817-6	Sequenc
37	71	7.2	257	9	US-09-893-817-22	Sequenc
38	70.5	7.1	262	11	US-09-764-875-1192	Sequenc
39	70.5	7.1	265	11	US-09-764-875-916	Sequenc
40	70.5	7.1	462	12	US-10-425-114-56491	Sequenc
41	69.5	7.0	120	12	US-10-424-599-155731	Sequenc
42	69.5	7.0	521	12	US-10-282-122A-48528	Sequenc
43	69.5	7.0	541	12	US-10-424-599-159290	Sequenc
44	69	7.0	457	11	US-09-826-509-579	Sequenc
45	69	7.0	490	12	US-10-425-114-53913	Sequenc

ALIGNMENTS

RESULT 1

US-09-147-801B-4
; Sequence 4, Application US/09147801B
; Publication No. US20030044960A1
; GENERAL INFORMATION:
; APPLICANT: Niklasson, Bo
; TITLE OF INVENTION: New Picorna Viruses, Vaccines and Diagnostic
; FILE REFERENCE: Niklasson 09/147801
; CURRENT APPLICATION NUMBER: US/09/147, 801B
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: PCT/SE97/01515
; PRIOR FILING DATE: 1997-09-09
; PRIOR APPLICATION NUMBER: SE 9603305-5
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 4
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Picornaviridae
US-09-147-801B-4

Query Match	100.0%	Score	990	DB	10	Length	179
Best Local Similarity	100.0%	Pred. No.	2.1e-103	Mismatches	0	Indels	0
Matches	179	Conservative	0	Mismatches	0	Indels	0
Qy	1	KDLMEIARMPVYKGERTEPGTNGYFOWSHTHSPINWVFDGGHLEDPNLL					
Db	1	KDLMEIARMPVYKGERTEPGTNGYFOWSHTHSPINWVFDGGHLEDPNLL					
Qy	61	YWRGSTVLKLTIVYASTFNKGLRMAFPIMMQGTQRKHKCLFMVCDTGLNNTF					
Db	61	YWRGSTVLKLTIVYASTFNKGLRMAFPIMMQGTQRKHKCLFMVCDTGLNNTF					
Qy	121	TWGNMTRGSGVIGWLRIIDLNLRLTYNSSSPNVCILQVKGNDGAKFMVPTT					
Db	121	TWGNMTRGSGVIGWLRIIDLNLRLTYNSSSPNVCILQVKGNDGAKFMVPTT					

```

481
Application US/10224999A
US20030171318A1
ATION:
riad Genetics, Inc.
orham, Scott
avitz, Kenton
oben, Adrian
NTION: Composition and Method for Treating Viral Infection
E: 5004.01
CATION NUMBER: US/10/224,999A
G DATE: 2003-03-03
TION NUMBER: US 60/313,695
DATE: 2001-08-20
ID NOS: 3484
entin version 3.1

man parechovirus 2
481
    41.8%; Score 413.5; DB 14; Length 2179;
ilarity 45.9%; Pred. NO. 5.8e-37;
Conservative 32; Mismatches 60; Indels 7; Gaps 3;
MEIARSPVYKGERTEPGGTN----GYQWSHTSHPINWVPDGGHILEDMENLFS 56
::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
VKISQIFSV-MASDTFSAHGIDQGYFKWANSDPQAIVHRNLVHLNLPFNKVF 419
::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
YNVRGSTVLKLTVYASTPNKGRLMAFFPIPMQGFTQRKKHKCLFMVCIDGLNNTFEM 116
::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
YSYFRGSLLIRLSVYASTFNRGRLN-GFFPNSSTDETSEIDNAIVTICIGSDNSFEI 478
::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
PYTWGWMMRTRGSVTCWLRIIDLVLNLTYNSSSPNAVNCILQVKMGNDAKFWPTTSN 176
::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
PYSFSTWMRKTHGKPGLGFQIEVLNLTNYSSPNEVYCIVGRKMGDQAKFFCPTGSL 538
::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

W 179
:
F 541

8
pplication US/10327481A
US20040001864A1
ATION:
ng, Andrew M.Q.
urman, Alison J.
udonnet, Jean-Christophe F.
ombard, Michel F.A.
ENTION: Vaccine against Foot-and-Mouth Disease
E: 454313-3178
CATION NUMBER: US/10/327,481A
G DATE: 2002-12-20
TION NUMBER: PCT/FR01/02042
DATE: 2001-06-27
TION NUMBER: FR 0008437
DATE: 2000-06-29
ID NOS: 41
entIn Ver. 2.1

ot-and-mouth disease virus
8
    8.8%; Score 87; DB 15; Length 1147;
ilarity 22.6%; Pred. No. 1.9;
Conservative 29; Mismatches 64; Indels 20; Gaps 6;

```

```

TION NUMBER: US 60/360.039
DATE: 2002-02-21
ID NOS: 47374
4

ccharomyces cerevisiae
044

      8.5%; Score 84; DB 15; Length 1108;
Identity 23.3%; Pred.No.3.9;
Conservative 25; Mismatches 70; Indels 40; Gaps 9;

GTNGYF-QWSHTSPINWVFDGGIHLEDMPNLNLFSscyNYWRGstVLKLTv-----72
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
SKNGYFGSOSVDCSP-----EEKIKLQECENMIVI-----GKTIFFELGHSYDGA 288

ASTFNKGLRMARFPIIMMQGTQRKKHKCLFMVCDTGLNNTFEMTIPYTWGNWMRPT-R 130
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
NVITYTQOQNVLDVPLALQNTFSKDGMCIAFFRDKSL---LASDLDFFRIARWVSPTFP 345

VIGWLRIDVNLRLTYNSS-----SPNAVNCILQVMGNDAKFWPPTTSNIW 179
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
IVG--LFDVFNDLRTNENILVPHFPNPGDHESI-----SSKNVYLDQTSNLSW 393

plication US/10135988
US20020176869A1
ATION:
NKHOUSER, ANN W
MERSON, SUZANNE U
JRCELL, ROBERT H
HONDT, ERIC
ATION: HEPATITIS A VIRUS VACCINES
3: 20264262US3
ATION NUMBER: US/10/135,988
3 DATE: 2002-04-29
TION NUMBER: 07/947,338
DATE: 1992-09-18
TION NUMBER: 08/397,232
DATE: 1995-03-10
ID NOS: 6
entIn Ver. 2.1

.D-TYPE HUMAN HEPATITIS A VIRUS, STRAIN HM-175

      8.4%; Score 83; DB 13; Length 2227;
Identity 26.7%; Pred.No.12;
Conservative 20; Mismatches 47; Indels 40; Gaps 9;

NLN-----LFSSC--YNYWRGstVLKLTvYASTFNKGLRMARFPP-----IMMOG- 93
|TNPDQKCI TALASICQMFCEWRGDLVDFQVFFPTKYHSGELLFCFVPGNELIDVSGI 392
|KKHKCLFMVCDI--GLNNTFEMTIPYTWGNWMRPT-----RG--SVIGWL 137
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
|QATTAPCAVMDITGVQSTLFRVP-----WISDTPYRVNRYTKSAHQKGEYTAIGKL 447

VNLRLTYNSSSPNAVNCILQVM 163
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
|CYNRLT-----SPSNVASHVRVN 469

plication US/10135988
US20020176869A1
ATION:

```


ATON: Hepatitis A virus sequence

publication US/10104966

ificial Sequence

ATION: Hepatitis A virus sequence

application US/09826660

ATION:
rdineau, Guy A.
elman, Steven J.
arva, Kenneth E.

```

Query Match      8.2%; Score 81.5; DB 9; Length 643;
Best Local Similarity 25.0%; Pred. No. 3.7;
Matches 36; Conservative 18; Mismatches 59; Indels 31;

QY      15  GERTPGG-- --NGVFQNSHTSPINWPDFDGI-----HLEDMP-NINLFS-----
Db      297  GRTNAPSGFASTN-----WFNWNPASFSALEAVIRPPHLLDPFEQLTIFSVLSF

QY      59  YNYWRGSTVCLKLTVYA--STFNKGRLRMAFFIMMGTQRKKHKKCLFMVCDGI
Db      353  MNYWGHRLSESTIRGSLSTSTGNTNTSINPTVLTQTSRDVYTESPA--GI

QY      114  FEMTIPYTWGNWRP---TRGSVI 134
Db      410  PVNGVPWAFENWRNPINLSLRGSL 433

```

```

US-09-826-660-23
/ Sequence 23, Application US/09826660
/ Patent No. US20010026940A1
/ GENERAL INFORMATION:
/ APPLICANT: Cardineau, Guy A.
/ APPLICANT: Stelman, Steven J.
/ APPLICANT: Narva, Kenneth E.
/ TITLE OF INVENTION: Plant-Optimized Genes Encoding Pesticidal Tox
/ FILE REFERENCE: MA-714XC2D1
/ CURRENT APPLICATION NUMBER: US/09/826,660
/ CURRENT FILING DATE: 2001-04-05
/ PRIOR APPLICATION NUMBER: 09/178,252
/ PRIOR FILING DATE: 1998-10-23
/ PRIOR APPLICATION NUMBER: 60/065,215
/ PRIOR FILING DATE: 1997-11-12
/ PRIOR APPLICATION NUMBER: 60/076,445
/ PRIOR FILING DATE: 1998-03-02
/ NUMBER OF SEQ ID NOS: 27
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 23
/ LENGTH: 1186
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Toxin encoded by synthetic B.t. gene
US-09-826-660-23

```

```

Query Match      8.2%; Score 81.5; DB 9; Length 1186;
Best Local Similarity 25.0%; Pred. No. 8.1;
Matches 36; Conservative 18; Mismatches 59; Indels 31;

15 GERTPEGG---TNGYFQWSHSTHSPINVFDDGI---HLEDP-NNLNFS---
QY

297 GRINAPSGFASTN---WFNNNAFSFAIRAAVTPRPPLDLPPEOLTFEVLNR
Db

```


us-09-147-801d-4.rapb

MTIPYTWGNWMP--TRGSVI 134
: : | | | | : :
NGVPWAREFNWRNPNSLRGSL 433

: April 23, 2004, 14:05:41
CS

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

cein search, using sw model

April 23, 2004, 13:57:19 ; Search time 20 Seconds
(without alignments)
860.914 Million cell updates/sec

JS-09-147-801D-4
990
1 KDLMEIARMPVYKGERTEP.....QVKGNDAKFMVPTTSNIW 179

3LOSUM62
3apop 10.0 , Gapext 0.5

283366 seqs, 96191526 residues

its satisfying chosen parameters: 283366

angth: 0
angth: 2000000000
Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

PIR 78: *
[: pir1: *
[: pir2: *
[: pir3: *
[: pir4: *

s the number of results predicted by chance to have a
er than or equal to the score of the result being printed,
ved by analysis of the total score distribution.

SUMMARIES

Query	Match	Length	DB	ID	Description
44.6	2180	2	A46182		polyprotein - echo
15.1	818	2	T08823		structural polypro
14.9	2303	1	GNNYTP		genome polyprotein
14.2	2303	2	S33554		genome polyprotein
13.9	2303	1	GNNYTM		genome polyprotein
13.5	929	2	A44048		capsid polyprotein
13.4	901	2	T03726		genome polyprotein
12.8	2301	1	GNNYTN		genome polyprotein
12.6	900	1	GNNYMY		genome polyprotein
12.5	426	2	S28374		genome polyprotein
12.4	2290	1	GNNYE		genome polyprotein
12.4	2292	1	GNNYED		genome polyprotein
12.4	2292	1	GNNYEB		genome polyprotein
12.4	2292	2	S35961		capsid polyprotein
12.4	2292	2	S55401		capsid polyprotein
10.3	811	2	T00036		capsid protein pre
9.9	2179	1	GNNYH4		genome polyprotein
9.5	2336	2	S37077		genome polyprotein
9.3	733	2	JQ1892		capsid protein - f
9.3	757	2	JQ1891		genome polyprotein - f
9.2	733	2	JQ1891		capsid protein - f
8.8	2333	1	GNNY2F		genome polyprotein
8.7	2332	1	GNNY4F		genome polyprotein
8.5	2332	1	GNNYF		genome polyprotein
8.5	1115	2	A47541		protein kinase IRE
8.4	874	2	T00491		capsid protein pre
8.4	1358	2	A03905		genome polyprotein
8.4	2227	1	GNNYHM		genome polyprotein
8.4	2227	1	GNNYMK		genome polyprotein

30	83	8.4	2227	1	GNNYHB	genome pc
31	82.5	8.3	637	2	JQ2039	polyprote
32	82.5	8.3	2206	1	GNNY27	genome pc
33	82.5	8.3	2209	1	GNNY3P	genome pc
34	82.5	8.3	3027	2	JQ1917	polyprote
35	82	8.3	852	1	GNNYHA	genome pc
36	82	8.3	2227	1	GNNYHR	genome po
37	82	8.3	2230	1	GNNYSA	genome po
38	81.5	8.2	1011	1	GNNYCI	genome po
39	81.5	8.2	2185	1	JQ2021	genome po
40	81.5	8.2	2207	1	GNNYIP	genome po
41	80.5	8.1	2201	1	GNNYA9	genome po
42	80	8.1	547	2	AG1813	hypotheti
43	79.5	8.0	2185	1	GNNYSV	genome po
44	79.5	8.0	2205	1	GNNY2W	genome po
45	73.5	8.0	2207	1	GNNY5P	genome po

ALIGNMENTS

RESULT 1
A46182
polyprotein - echovirus 22
N:Contains: protein 2A; protein 2B; protein 2C; protein 3A; protein
C:Species: echovirus 22
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 08-0
C:Accession: A46182
R:Hyypia, T.; Horsnell, C.; Maaronen, M.; Khan, M.; Kalkkinen, N.; A
Proc. Natl. Acad. Sci. U.S.A. 89, 8847-8851, 1992
A:Title: A distinct picornavirus group identified by sequence analys
A:Reference number: A46182; MUID:92409614; PMID:1528901
A:Accession: A46182
A:Status: preliminary
A:Molecule type: genomic RNA; protein
A:Residues: 1-2180 <HY>
A:Cross-references: GB:S45208; GB:L00675; NID:g256078; PIDN:AAB23363
A:Experimental source: strain Harris
A:Note: sequence extracted from NCB1 backbone (NCBIN:114262, NCBIP:1
C:Keywords: polyprotein

Query Match 44.6%; Score 442; DB 2; Length 2180;
Best Local Similarity 48.6%; Pred. No. 4e-35;
Matches 89; Conservative 33; Mismatches 55; Indels 6;
Oy 2 DLMEIARMPVYKGERTEP-----GTNGYFQWSTHSPINWVFDGGIHLSDMP
Db 360 DLVIAQLFSV-MADSTTPSENHGVDKGYFKWSATTAPQSIVHRNIYVLRFP
Oy 57 SCYNWVRGSLVLTVAFTNKGRMRMAFPPIWQQGTQRKKHCKLFWVCDIGL
Db 419 NSYSYFRGSLVLRSLVSVAFTNCRGLRMGFPFNATDTSTLDNAIYICDIGS
Oy 117 TIPTWGNWRPPTGSGVIGWLRIDVLRNLTYNSSSPNAVNCILQVKGNDAKFM
Db 479 TIPYFSFTWKRKTGHPGLFQIEVLNRLTYNSSSPSEVYCIQVQKMGQDARFF
Oy 177 IVW 179
Db 539 VTF 541

RESULT 2
T08823
structural polyprotein - Rhopalosiphum padi virus
C:Species: Rhopalosiphum padi virus
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Ju
C:Accession: T08823
R:Moore, J.S.; Domier, L.L.; McCoppin, N.K.; D'Arcy, C.J.; Jin, H.
Virology 243, 54-65, 1998
A:Title: Nucleotide sequence analysis shows that Rhopalosiphum padi
A:Reference number: 216481; MUID:98201645; PMID:9527915
A:Accession: T08823

ated from GB/EMBL/DBJ
8 <WOO>
es: EMBL:AF022937; NID:G2911298; PIDN:AA95510.1; PID:G2911300
of translation initiation could not be determined from the nucleotide s
protein
15.1%; Score 149; DB 2; Length 818;
ilarity 20.5%; Pred. No. 1.2e-06;
Conservative 36; Mismatches 83; Indels 52; Gaps 4;
ARMPSVYKGERTE-----PGGTNGYFQWSTHSPINWVFGGHH 45
EHLQSLFRDSDMSIDVARTYNYVSFPNWKNGPGTLYSHVSPSTWFTVGIT 391
DMPNINLFSSCYNWRSSTVLKLTYYASTFNKGRMAFFIMMQGTQRKHKC---L 102
SIPHLFYAASNFVLRGGMNKLKFKVTKFHSGRVILYVPGFGGTLPTNFETDANY 451
VCDIGLNTTFEMTIPYTWGNWRTRGSRVIGWLR-----IDVL 142
VVDLRSDPDEFFNPFYV-----ATVPLHVNSTPMTAFSQIHACKRSIVVEVL 501
LTYNSSPNAVCILQVKGNDKAFMVPPTNSI 177
LVNTSTVSDTIEVLVEVCAEDIEFAIPVPSL 536
in - murine poliovirus (strain GDVII)
able proteinase (EC 3.4.-.-); protein 1A; protein 1B; protein 1C; protei
s poliovirus, Theiler's encephalomyelitis virus
990 #sequence_revision 31-Mar-1990 #text_change 16-Jul-1999
193
Borkowski, J.; Calenoff, M.; Oh, C.K.; Ostrowski, B.; Lipton, H.L.
12, 1988
s into Theiler's virus neurovirulence based on a genomic comparison of t
er: A29193; MUID:88265847; PMID:2838951
193
genomic RNA
03 <PEV>
es: GB:M20562; NID:G335221; PIDN:AAA96329.1; PID:G1256538
oot-and-mouth disease virus genome polyprotein
protein; core protein; genome-linked protein; hydrolase; polyprotein; P
eader peptide #status predicted <LDP>
t: protein 1A (coat protein VP4) #status predicted <VP4>
t: protein 1B (coat protein VP2) #status predicted <VP2>
t: protein 1C (coat protein VP3) #status predicted <VP3>
t: protein 1D (coat protein VP1) #status predicted <VP1>
ct: protein 2A (core protein P2-3b) #status predicted <P2A>
uct: protein 2B (core protein P2-5b) #status predicted <P2B>
uct: protein 2C (core protein P2-X) #status predicted <P2C>
uct: protein 3A (genome-linked protein VPg) #status predicted <P3A>
uct: protein 3B (genome-linked protein VPg) #status predicted <P3B>
uct: protein 3C (probable proteinase) #status predicted <P3C>
uct: protein 3D (probable RNA-directed RNA polymerase) #status predicted
14.9%; Score 148; DB 1; Length 2303;
ilarity 25.5%; Pred. No. 5.2e-06;
Conservative 34; Mismatches 87; Indels 16; Gaps 5;
MEIARMPSVYKGERTEPGGTNGYFQWSTHSPINWVFDGGIHLDPNINLSSC-- 58
LELCKLPTFLGNPSTD---NKRYPYFSATNSVPATSLVDYQVALSCSTANSMLAAVA 513
YNYWRGSLTVLKLTYASTFNKGRMAFFP---IMMQGTQRKHKLPMVCDIGLNTTF 114
FNQYRGSLNLFVFTGAAWVGKFKFLIATPPGAGKPTTRDQAMQATYAIWDLGNSSF 573
TIP-----YTWGNWMPTRGSRVIGWLRIDVLNRLTYNSSSPNAVNCILQVKGNDK 168

Db 574 NETAPEISPTHYRQTSYTSPTITSVDGWTWQLPLTYPSPGTPHSDILTLVS
QY 169 FMVP 172
Db 634 LRMP 637
RESULT 4
S13554
genome polyprotein - murine poliovirus
C:Species: murine poliovirus, Theiler's encephalomyelitis virus
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 26-f
C:Accession: S13554
R:Law, K.M.; Brown, T.D.K.
Nucleic Acids Res. 18, 6707-6708, 1990
A:Title: The complete nucleotide sequence of the GDVII strain of The
A:Reference number: S13554; MUID:91067481; PMID:2251141
A:Accession: S13554
A:Status: preliminary; translation not shown
A:Molecule type: mRNA
A:Residues: 1-2303 <LAW>
A:Cross-references: EMBL:X56019; NID:G62039; PIDN:CRA39496.1; PID:G6
C:Superfamily: foot-and-mouth disease virus genome polyprotein
Query Match 14.2%; Score 141; DB 2; Length 2303;
Best Local Similarity 25.0%; Pred. No. 2.6e-05;
Matches 46; Conservative 34; Mismatches 88; Indels 16;
QY 2 DLMEIARMPSVYKGERTEPGGTNGYFQWSTHSPINWVFDGGIHLDPNINLNI
Db 457 DLELCKLPTFLGNPSTD---NKRYPYFSATNSVPATSLVDYQVALSCSTANS
QY 59 ---YNYWRGSLTVLKLTYASTFNKGRMAFFP---IMMQGTQRKHKLPMVCDI
Db 514 RNFNQYRGSLNLFVFTGAAWVGKFKFLIATPPGAGKPTTRDQAMQATYAIWDI
QY 115 EMTIP-----YTWGNWMPTRGSRVIGWLRIDVLNRLTYNSSSPNAVNCILOVK
Db 574 NETAPEISPTHYRQTSYTSPTITSVDGWTWQLPLTYPSPGTPHSDILTLVS
QY 169 FMVP 172
Db 634 LRMP 637
RESULT 5
GNMYTM
genome polyprotein - murine poliovirus (strain Bean 8386)
N:Contains: probable proteinase (EC 3.4.-.-); protein 1A; protein 1B
C:Species: murine poliovirus, Theiler's encephalomyelitis virus
C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 16-J
C:Accession: A29535
R:Pevear, D.C.; Calenoff, M.; Rozhon, E.; Lipton, H.L.
J. Virol. 61, 1507-1516, 1987
A:Title: Analysis of the complete nucleotide sequence of the picorna
A:Reference number: A29535; MUID:87198877; PMID:3033278
A:Accession: A29535
A:Molecule type: Genomic RNA
A:Residues: 1-2303 <PEV>
A:Cross-references: GB:M16020; NID:G335239; PIDN:AAA47930.1; PID:G33
C:Superfamily: foot-and-mouth disease virus genome polyprotein
C:Keywords: coat protein; core protein; genome-linked protein; hydro
F:1-76/Domain: leader peptide #status predicted <LDP>
F:77-147/Product: protein 1A (coat protein VP4) #status predicted <V
F:148-414/Product: protein 1B (coat protein VP2) #status predicted <V
F:415-646/Product: protein 1C (coat protein VP3) #status predicted <V
F:647-922/Product: protein 1D (coat protein VP1) #status predicted <V
F:923-1061/Product: protein 2A (core protein P2-3b) #status predicte
F:1062-1191/Product: protein 2B (core protein P2-5b) #status predicte
F:1192-1517/Product: protein 2C (core protein P2-X) #status predicte
F:1518-1605/Product: protein 3A (protein P3-1b) #status predicted <P
F:1606-1625/Product: protein 3B (genome-linked protein VPg) #status]

```

ct: protein 3C (probable proteinase) #status predicted <P3C>
ct: protein 3D (probable RNA-directed RNA polymerase) #status predicted<P3C>

          13.9%; Score 138; DB 1; Length 2303;
          larity 26.1%; Pred. No. 5.1e-05;
Conservative 33; Mismatches 87; Indels 16; Gaps 6;

EIARMPSSVYKGBRTEPGGTNGYFQWSHTHS-PINWVPDGGIHL--EDMPN--LNLFSS 56
ELCKLTFTFLGNPNT---NNKRYPYFSATNSVPATSMVDYQVALSCSOMANSLAAVA 513
NYWRGTSVLKLTVYASTENKGRLMARFP--IMMQGTQRKKHCLFMUCDIGLANTFF 114
NQVRGSINLFLVFETGAMVKGKGLIAYTPPGAGKPTTRDAQMSTTAVMDUGLINSF 573
TP-----YTWNMRPRGSGVIGWLRIDVLNLTYNSSPNNAVNCILQVKNMGDAK 168
APFISPTHYRQTSYSPIITSDVGWVTWKLTPLTYPSGTPTNSDILTIVSAGDDFT 633

P 172
|
P 637
|

n - Vilyuisk virus (strain V-1) (fragment)
93 sk virus
93 #sequence_revision 11-Feb-1993 #text_change 07-Feb-1997
48
: Strom, T.; Lipton, H.L.
-472, 1992
de sequence identifies Vilyuisk virus as a divergent Theiler's virus.
r: A44046; MUID:93033144; PMID:1413519
48
genomic RNA
<PRI>
S: GB:M94868
ct-and-mouth disease virus genome polyprotein
rotein

          13.5%; Score 133.5; DB 2; Length 929;
          larity 21.2%; Pred. No. 4.9e-05;
Conservative 39; Mismatches 78; Indels 35; Gaps 6;

EIARMPSSVYKGBRTEPGGTNGYFQWSHTSPINWVPDGGIHLDMENLNFSSC--- 58
ELCKLT-T-FLGNLS-----NDRVPFF-FTATNSVPTESIVEYQVTLSCSM 503
-----YYNRGSTVLKLVYASTFNKGRLMARFP--IMMQGTQRKKHKCLFMV 105
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
MLASVARNEQNVRGSLNFLFVFTGSAITKGFLLIAVTPPGAGKPTTRDAQXSTVAI 563
GLNTEFEMTIP-----YTWNMRPRRGSGVIGWLRIDVLNLTYNSSPNNAVNCIL 159
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
GLNSSLNFTVPFIISPESHYRQTSYSTSPSIAAVDGMLTWQLTPLTFPANVPSSDILT 623
MGNDAKFMVP 172
||| . |
AGNDFLTRMP 636

n - Drosophila C virus (strain EB) (fragment)
hila C virus
EB
99 #sequence_revision 23-Apr-1999 #text_change 17-Nov-2000
26
Christian, P.D.
, 191-203, 1998
r genome organization of the insect picorna-like virus Drosophila C vir
r: Z15031; MUID:98120840; PMID:9460942
```

```
IGLNNTFEWCLP-----YTWGWMRRTQSGVIGLRLDVIWNRIITYNSSSPNAVNCIL 159
:|::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
:LGNSSEVFAPFISPHFYEQTSYTSATIASVDGVWTWQLTSLTPESGAFVNSDIIT 624
XMGNDAKFMVP 172
:|::||:
SAGDDFTLRMP 637
```

in - Mengo virus (strain 37A) (fragment)
protein VP1; coat protein VP2; coat protein VP3; coat protein VP4
virus

erson, K.; Luo, M.; Bond, C. W.
7-345, 1992
ar and structural basis of hemagglutination in mengovirus.
er: A43379; MUID:92410611; PMID:326807
379

```

genomic RNA
0 <MAN>
es: GB:M88547
translation of the VP1 nucleotide sequence is given
cot-and-mouth disease virus genome polyprotein
protein; polyprotein
leader peptide #status predicted <LDP>
t: coat protein VP4 #status predicted <VP4>
t: coat protein VP2 #status predicted <VP2>
t: coat protein VP3 #status predicted <VP3>
t: coat protein VP1 #status predicted <VP1>

```

12.6%; Score 124.5; DB 1; Length 900;
 ilarity 23.2%; Pred. No. 0.00037;
 Conservative 35; Mismatches 74; Indels 43; Gaps 10;

[illegible]

NY--WRGSTVLKLTYYASTFNKGRURMAFFP-----INMQGTQRKKHKCLFMV 105
| : |||| | : ||| : || :
NEMOVRGCTVTMEYFETCTAMNKGVNTTVMRDRQAGVATPCATCVCSTCTC

[illegible]

.GLNSSYSFVTF-----ISPTFRMVGTDLVNITNADGWVTVQLTPLTFPGCPTS 597
 .
 .CILQVMGNDAKFMVP 172
 .
 .

ILTMVSAGKDFSLKMP 615

```

in - cricket paralysis virus (fragment)
ac paralysis virus
333 #sequence_revision 17-Apr-1993 #text_change 08-Oct-1999
374
Illin, J.S.K.; Stanway, G.; Almond, J.W.; Moore, N.F.
1-344, 1987
of the genome of cricket paralysis virus: sequence of the 3' end.

```

374 genomic RNA
5<KIN>
es: ENBL:M21938; NID:G323280; PID:AAA42885.1; PID:G323281
protein

```
12.5%; Score 123.5; DB 2; Length 426;
Similarity 22.7%; Pred. No. 0.00019;
```

[illegible]

RESULT 11

genome polyprotein - encephalomyocarditis virus
N:Contains: coat protein VP1; coat protein VP2; coat protein VP3; co
EC 3.4.-); RNA-directed RNA polymerase (EC 2.7.7.48)
C:Species: encephalomyocarditis virus, EMCV
A:Note: host Homo sapiens (man)
C:Date: 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change 16-J
C:Accession: A03906; JN0383
R:Palmenberg, A.C.; Kirby, E.M.; Janda, M.R.; Drake, N.L.; Duke, G.M
Nucleic Acids Res. 12, 2969-2985, 1984
A:Title: The nucleotide and deduced amino acid sequences of the ence
A:Reference number: A03906; MUID:84169586; PMID:6324136
A:Accession: A03906
A:Molecule type: genomic RNA
A:Residues: 1-2290 <PRL>
A:Cross-references: GB:X00463; NID:961034; PTDN:CA825152.1; PID:9610
R:Petrov, N.A.; Chizhikov, V.E.; Blinov, V.M.; Karginov, V.A.; Mikry
Bioorg. Khim. 10, 274-279, 1984
A:Title: Nucleotide sequence of the 3'-terminus of encephalomyocardi
A:Reference number: JN0383; MUID:85022788; PMID:6091680
A:Accession: JN0383

A,Molecule type: genomic RNA
A,Residues: 1337-1396,'L',1398-1517,'A',1519-1536,'E',1538-1556,'S',
A,Cross-references: GB:M54935
A,Note: the authors translated the codon CAU for residue 713 as Thr
C,Superfamily: foot-and-mouth disease virus genome polyprotein
C,Keywords: coat protein; core protein; genome-linked protein; hydro
F,1-67/Domain: leader peptide #status predicted <Ldp>
F,68-136/Product: coat protein VP2 #status predicted <VP4>
F,137-391/Product: coat protein VP4 #status predicted <VP2>
F,392-625/Product: coat protein VP3 #status predicted <VP3>
F,623-910/Product: coat protein VP1 #status predicted <VP1>
F,911-1056/Product: core protein P2-A #status predicted <P2A>
F,1057-1192/Product: core protein P2-B #status predicted <P2B>
F,1193-1517/Product: core protein P2-C #status predicted <P2C>
F,1518-1605/Product: core protein P3-A #status predicted <P3A>
F,1606-1625/Product: genome-linked protein VPg #status predicted <VPg>
F,1626-1830/Product: proteinase #status predicted <PTS>
F,1831-2290/Product: RNA-directed RNA polymerase #status predicted <P>

Query Match	12.4%	Score 122.5;	DB 1;	Length 2290;
Best Local Similarity	20.8%;	Pred.	No. 0.0017;	
Matches	43;	Conservative	34;	Mismatches 69; Indels 61;
QY	1	KDLMEIAMPVYKGERTEPGGTNGYFQWGHHSPIINWFDGGIHLEDMENLNLI		
Ddb	433	KDFLEIAQIPT-FIGNKI-PNAV-PYEASNT-----AVKTOPLATYQV	:	:
QY	59	-----YNVRGSGTVLKLTVYASTFNKGRLSMAFP-----IM	:	:
Ddb	480	LANTFLAALSENFAQYRGSLVYVFVTGTAMMKGFLLIAYTPPGAGKPTSRDQA	:	:
QY	97	KHKHCLFVWCIDIGLNNFTFMTIPYTWNMWRPRTGSVI-----CWLRII	:	:
Ddb	538	-----YAIDLGLNSSYSFTVP-----ISPTHFRMVGTQVNITNADGWTVI	:	:
QY	146	TNVSSTSSNAVNCITLOVFWGNDAKEWVD	172	

PPGCPTSAGKILTMWSAGKDFSLKMP 613

in - encephalomyocarditis virus (strain EMC-D, diabetogenic)
protein VP1; coat protein VP2; coat protein VP3; coat protein VP4; core
-directed RNA polymerase (EC 2.7.7.48)

halomyocarditis virus, EMCV
o sapiens (man)

990 #sequence_revision 31-Mar-1990 #text_change 13-Mar-1998

473; A60498
H.M.; Yoon, J.W.

2-287, 1989
differences between the diabetogenic and nondiabetogenic variants of en

er: A94395; MUID:89243189; PMID:2541543

473
Genomic RNA

ors translated the codon ATG for residue 1079 as Asn and GAC for residue

H.M.; Yoon, J.W.

320, 1989
ar identification of diabetogenic viral gene.

er: A60498; MUID:89137787; PMID:2537245

498
Genomic RNA

22 'D' 1524-2292 <BA2>
oot-and-mouth disease virus genome polyprotein

protein; core protein; genome-linked protein; hydrolase; nucleotidyltra

sader peptide #status predicted <LDP>

t: coat protein VP4 #status predicted <VP4>

t: coat protein VP2 #status predicted <VP2>

t: coat protein VP3 #status predicted <VP3>

t: coat protein VP1 #status predicted <VP1>

ct: core protein P2-A #status predicted <P2A>

ct: core protein P2-B #status predicted <P2B>

ct: core protein P2-C #status predicted <P2C>

ct: core protein P3-A #status predicted <P3A>

ct: genome-linked protein VPg #status predicted <VPG>

ct: proteinase #status predicted <PTS>

ct: RNA-directed RNA polymerase #status predicted <RDP>

12.4%; Score 122.5; DB 1; Length 2292;

ilarity 21.3%; Pred. No. 0.0017;

Conservative 33; Mismatches 69; Indels 61; Gaps 9;

MEIARMPVYKGERTEPGTNGYFQWSHSPINWVFDGGLHLEMDPNLNFSSC-- 58

FLBIAQIPT-FIGNKI-PNAV-PYIASNT-----AVKTQPLATYQVTLSCSC 481

-----YNYWRGSLVTLKTYASTFNKGRIRMAFFP-----IMOGTQR 96

VTFLAALSRNFAQYRGLSVTFVTGTAMMKGRFLIAYTPPGAGKPTRDQAMQAT-- 539

IKCLFVCDIGLNTTEMTIPYTWGNMRPTR-----GSVIGWLRIIDLNLRL 145

-----YAIWDLGLNSSYSFTVPF-----ISPTFRMVGTDOVNTVDGWTWVQLPL 588

YSSSPNAVNCILQVKNDAKFMVP 172

PPGCPTSAGKILTMWSAGKDFSLKMP 615

in - encephalomyocarditis virus (strain EMC-B, nondiabetogenic)

protein VP1; coat protein VP2; coat protein VP3; coat protein VP4; core

-directed RNA polymerase (EC 2.7.7.48)

halomyocarditis virus, EMCV
o sapiens (man)

990 #sequence_revision 31-Mar-1990 #text_change 13-Mar-1998

173; B60498

R.Bae, Y.S.; Eun, H.M.; Yoon, J.W.

Virology 170, 282-287, 1989

A>Title: Genomic differences between the diabetogenic and nondiabeto

A;Reference number: A94395; MUID:89243189; PMID:2541543

A;Accession: B31473

A;Molecule type: genomic RNA

A;Residues: 1-2292 <BA2>

A;Note: the authors translated the codon ATG for residue 1079 as Asn

R.Bae, Y.S.; Eun, H.M.; Yoon, J.W.

Diabetes 38, 316-320, 1989

A>Title: Molecular identification of diabetogenic viral gene.

A;Reference number: A60498; MUID:89137787; PMID:2537245

A;Accession: B60498

A;Molecule type: genomic RNA

A;Residues: 1-1517-2292 <BA2>

C;Superfamily: foot-and-mouth disease virus genome polyprotein

C;Keywords: coat protein; core protein; genome-linked protein; hydrc

F1-67/Domain: leader peptide #status predicted <LDP>

F168-137/Product: coat protein VP4 #status predicted <VP4>

F138-393/Product: coat protein VP2 #status predicted <VP2>

F1394-624/Product: coat protein VP3 #status predicted <VP3>

F1625-901/Product: coat protein VP1 #status predicted <VP1>

F1902-1058/Product: core protein P2-A #status predicted <P2A>

F1053-1194/Product: core protein P2-B #status predicted <P2B>

F1195-1519/Product: core protein P2-C #status predicted <P2C>

F1520-1607/Product: core protein P3-A #status predicted <P3A>

F1608-1627/Product: genome-linked protein VPg #status predicted <VPG>

F1628-1832/Product: proteinase #status predicted <PTS>

F1833-2292/Product: RNA-directed RNA polymerase #status predicted <

Query Match 12.4%; Score 122.5; DB 1; Length 2292;

Best Local Similarity 21.3%; Pred. No. 0.0017;

Matches 44; Conservative 33; Mismatches 69; Indels 61;

QY 1 KDLMEIARMPVYKGERTEPGTNGYFQWSHSPINWVFDGGLHLEMDPNLNL

DB 435 KDFLEIAQIPT-FIGNKI-PNAV-PYIASNT-----AVKTQPLATYQV

QY 59 -----YNYWRGSLVTLKTYASTFNKGRIRMAFFP-----IM

DB 482 LANTFLAALSRNFAQYRGLSVTFVTGTAMMKGRFLIAYTPPGAGKPTRDQA

QY 97 KKHKCLFVCDIGLNTTEMTIPYTWGNMRPTR-----GSVIGWLRI

DB 540 -----YAIWDLGLNSSYSFTVPF-----ISPTFRMVGTDOVNTVDGWTWV

QY 146 TYSSSPNAVNCILQVKNDAKFMVP 172

DB 589 TYPGCPTSAGKILTMWSAGKDFSLKMP 615

RESULT 14

S35961

capsid polyprotein precursor - encephalomyocarditis virus

C;Species: encephalomyocarditis virus, EMCV

C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 26-A

C;Accession: S35961

R;Zimmermann, A.; Nelsen-Slax, B.; Kruppenbacher, J.P.; Eggers, H.J.

submitted to the EMBL Data Library, July 1993

A;Description: The complete nucleotide sequence and construction of

A;Reference number: S35961

A;Accession: S35961

A;Status: preliminary

A;Molecule type: genomic RNA

A;Residues: 1-2292 <ZIM>

A;Cross-references: EMBL:X74312; NID:G396509; PIDN:CAA52361.1; PID:G

C;Superfamily: foot-and-mouth disease virus genome polyprotein

C;Keywords: polyprotein

Query Match 12.4%; Score 122.5; DB 2; Length 2292;

Best Local Similarity 20.8%; Pred. No. 0.0017;

Matches 43; Conservative 34; Mismatches 69; Indels 61;

[illegible]

in precursor - encephalomyocarditis virus
nalomyocarditis virus, EMCV
995 #sequence_revision 03-Nov-1995 #text_change 26-Aug-1999
401
.: Zimmermann, A.; Wickert, S.; Arnold, G.; Botta, A.; Eggers, H.; Krupp
EMBL Data Library, May 1995
analysis of sequence and pathogenic properties of two variants of enceph
er: S55401
401
inary
Genomic RNA
32 <NEL>
es: EMBL:X87335; NID:g854638; PIDN:CAA60776.1; PID:g854639
ot-and-mouth disease virus genome polypeptide
protein

```

12.4%; Score 122.5; DB 2; Length 2292;
ilarity 21.3%; Pred.No.0.0017;
Conservative 33; MisMatches 69; Indels 61; Gaps 9
MEIARMPGVYKGERTEPGGTNGCYFOWSHTHSPINWVFDGGIHLEDPNLIFFSSC-- 58
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
TLEIAQIPT-FIGNKI-PNAV-PYIASNT-----AVKTOPLATYQVTLSGSC 481
-----YNWRGSTVLKLTYYASTNKGBLRMAFTP-----IMMQGTQR 96
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
VTFLAALSRNFAYRGSLVYTFVFTGMWKGLFIATPPGAGKPTSRDQAMQAT-- 539
IKCLFMVCIDIGLNNTEMTIPYTWGNWMRPTR-----GSVIGMLRIDVLNRL 145
-----YAIDLGLNSSYSFTVPF-----ISPTHRMVGTQDVNIITVDGWTVVQLTPL 588
ISSSNPANNCCILOVKMGNDAKFMVP 172
PGCGPTSAKILTWVSAGKDfSLKNP 615

```

: April 23, 2004, 14:00:10

02

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

tein search, using sw model

April 23, 2004, 13:57:18 ; Search time 18 seconds
(without alignments)

517.809 Million cell updates/sec

JS-09-147-801D-4

990

1 KDLMEIARMPVYKGERTEP.....QVKMGNDKFMVPTTSNIW 179

BLOSUM62

Gapop 10.0 , Gapext 0.5

141681 segs, 52070155 residues

hits satisfying chosen parameters: 141681

length: 0

length: 2000000000

Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

SwissProt_42:*

is the number of results predicted by chance to have a
per than or equal to the score of the result being printed,
ived by analysis of the total score distribution.

SUMMARIES

Query	Match	Length	DB	ID	Description
44.6	2180	1	POLG_EC22H	Q66578	e genome po
43.1	2188	1	POLG_EC23C	Q9yld8	e genome po
41.8	2179	1	POLG_EC23W	O73556	e genome po
14.9	2303	1	POLG_TMEVG	P08545	t genome po
13.9	2303	1	POLG_TMEVB	P08544	t genome po
12.8	2301	1	POLG_TMEVD	P13899	t genome po
12.6	834	1	POLG_ENMG3	P12296	mengo encep
12.6	901	1	POLG_ENMG3	P32540	mengo encep
12.5	426	1	POLG_CRPV	P13418	cricket par
12.4	2290	1	POLG_EMCV	P03304	e genome po
12.4	2292	1	POLG_EMCVB	P17593	e genome po
12.4	2292	1	POLG_EMCVD	P17594	e genome po
10.6	855	1	POLG_HRV3	O82081	human rhino
9.9	2179	1	POLG_HRV14	P03303	h genome po
9.8	2184	1	POLG_CXB6S	Q9ql88	c genome po
9.5	2336	1	POLG_FMDVZ	P49303	f genome po
9.3	2184	1	POLG_ECO1F	O91734	e genome po
8.8	2333	1	POLG_FMDV1	P03306	f genome po
8.7	2332	1	POLG_FMDVA	P03308	f genome po
8.5	2332	1	POLG_FMDVO	P03305	f genome po
8.5	1115	1	IRE1_YEAST	P32361	saccharomyc
8.4	2226	1	POLG_HPAV2	P26581	hepatitis a
8.4	2226	1	POLG_HPAV4	P26580	hepatitis a
8.4	2227	1	POLG_HPAVH	P08617	hepatitis a
8.4	2227	1	POLG_HPAVM	P13901	hepatitis a
8.3	2206	1	POLG_POL32	P06209	pollowirus
8.3	2209	1	POLG_POL1S	P03301	p genome po
8.3	3027	1	POLG_PYFV1	Q05057	parsnip yel
8.3	852	1	POLG_HPAVC	P06442	hepatitis a
8.3	2168	1	POLG_PEV9U	O41174	p genome po
8.3	2226	1	POLG_HPAVB	P26582	hepatitis a
8.3	2227	1	POLG_HPAVL	P06441	hepatitis a
8.3	2230	1	POLG_HPAVS	P14553	simian hepa

ALIGNMENTS

RESULT 1

ID	POLG_EC22H	STANDARD;	PRT;	2180 AA.
AC	Q66578; Q30062;			
DT	15-JUL-1999 (Rel. 38, Created)			
DT	15-JUL-1999 (Rel. 38, Last sequence update)			
DE	Genome polyprotein [Contains: Coat protein VP0 (PIAB); Coat pro			
DE	VP3 (PIC); Coat protein VP1 (PID); Core protein P2A; Core prote			
DE	Core protein P2C; Core protein P3A; Genome-linked protein VPG (
DE	Picornain 3C (EC 3.4.22.28) (Protease 3C) (P3C); RNA-directed R			
DE	Polymerase (EC 2.7.7.48) (P3D)].			
OS	Echovirus 22 (strain Harris) (Human parechovirus 1).			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Picornavi			
OC	Parechovirus.			
OX	NCBI_TaxID=103911;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 29-61; 67-80; 290-297 AND 5			
RX	MEDLINE=92409614; PubMed=1528901;			
RA	Hyypiae T., Horsnell C., Maaronen M., Khan M., Kalkkinen N.,			
RA	Auvinen P., Kinnunen L., Stanway G.;			
RT	"A distinct picornavirus group identified by sequence analysis.			
RL	Proc. Natl. Acad. Sci. U.S.A. 89:8847-8851(1992).			
RN	[2]			
RP	SEQUENCE OF 6-22; 29-61; 67-80 AND 118-131.			
RX	MEDLINE=9505053; PubMed=7966616;			
RA	Stanway G., Kalkkinen N., Roivainen M., Ghazi F., Khan M., Smyt			
RA	Meurman O., Hyypiae T.;			
RT	"Molecular and biological characteristics of echovirus 22, a			
RT	representative of a new picornavirus group."			
RL	J. Virol. 68:8232-8238(1994).			
CC	!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES			
CC	AT CERTAIN Q/G SITES IN THE POLYPROTEIN. IT IS A CYSTEINE			
CC	PROTEASE.			
CC	!- CATALYTIC ACTIVITY: Selective cleavage of Gln-[Gly bond in			
CC	poliiovirus polyprotein. In other picornavirus reactions Glu			
CC	substituted for Gln, and Ser or Thr for Gly.			
CC	!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphospha			
CC	{RNA} (N).			
CC	!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNI			
CC	EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP0,			
CC	AND VP3.			
CC	!- PTM: Specific enzymatic cleavages in vivo yield mature prot.			
CC	ALL CLEAVAGES ARE CATALYZED BY P3C.			
CC	!- PTM: THE N-TERMINUS OF VP0 IS BLOCKED.			
CC	!- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a co			
CC	between the Swiss Institute of Bioinformatics and the EMBL o			
CC	the European Bioinformatics Institute. There are no restricti			
CC	use by non-profit institutions as long as its content is			
CC	modified and this statement is not removed. Usage by and for			
CC	entities requires a license agreement (See http://www.isb-sib.ch)			
CC	or send an email to license@isb-sib.ch .			
CC	-----			
CC	EMBL; L02971; AAA72291.1; -.			

Virus Res. 56:217-223(1998).

-!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN Q/G SITES IN THE POLYPROTEIN. IT IS A CYSTEINE PROTEASE.

-!- CATALYTIC ACTIVITY: Selective cleavage of Gln-Gly bond in poliovirus polyprotein. In other picornavirus reactions Gln substituted for Gln, and Ser or Thr for Gly.

-!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate (RNA) (N).

-!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP0 AND VP3.

-!- PTM: Specific enzymatic cleavages in vivo yield mature proteins. ALL CLEAVAGES ARE CATALYZED BY P3C.

-!- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL European Bioinformatics Institute. There are no restrictions on use by non-profit institutions as long as its content is modified and this statement is not removed. Usage by and for entities requires a license agreement (See <http://www.isb-sib.ch>) or send an email to license@isb-sib.ch.

EMBL; AF055846; AAC79756.1; --
MEROPS: C03_023; --
InterPro; IPR004004; Calici_pol_hel.
InterPro; IPR003003; Cys_Ser_trypsin.
InterPro; IPR007053; NC.
DR InterPro; IPR000605; RNA_helicase.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR001205; RNA_pol_P3D.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR InterPro; IPR008975; Viral_cap_coat.
Pfam; PF04970; NC; 1.
Pfam; PF00680; RNA_dep_RNA_pol; 1.
Pfam; PF00910; RNA_helicase; 1.
PRINTS; PR00918; CALICIVIRUSNS.
KW Polyprotein; Coat protein; Core protein; Transferase;
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease;
FT CHAIN 1 290 COAT PROTEIN VP0 (BY SIMILARITY).
FT CHAIN 291 549 COAT PROTEIN VP3 (BY SIMILARITY).
FT CHAIN 550 784 COAT PROTEIN VP1 (BY SIMILARITY).
FT CHAIN 785 931 CORE PROTEIN 2A (BY SIMILARITY).
FT CHAIN 932 1053 CORE PROTEIN P2B (BY SIMILARITY).
FT CHAIN 1054 1382 CORE PROTEIN P2C (BY SIMILARITY).
FT CHAIN 1383 1499 CORE PROTEIN P3A (BY SIMILARITY).
FT CHAIN 1500 1519 GENOME-LINKED PROTEIN VPG (BY SIMILARITY).
FT CHAIN 1520 1719 PICORNAIN 3C (BY SIMILARITY).
FT CHAIN 1720 2188 RNA-DIRECTED RNA POLYMERASE (BY SIMILARITY).
SITE 772 774 CELL ATTACHMENT SITE (POTENTIAL).
FT ACT_SITE 1678 1678 PROTEASE 3C (POTENTIAL).
FT ACT_SITE 1696 1696 PROTEASE 3C (POTENTIAL).
SQ SEQUENCE 2188 AA; 246602 MW; 02CC77D0A5ED3D93 CRC64;

Query Match 43.1%; Score 427; DB 1; Length 2188;
Best Local Similarity 46.4%; Pred. No. 2, 2e-35;
Matches 83; Conservative 32; Mismatches 58; Indels 6;

QY 2 DLMEIARMPSVYKGERTEPGTN-----GYFQWSHTHSPINVPDGGIHLEDMP
Db 367 DMIIIAQLFSV-MSDTTTPSTSGIDDLGLVDWSATYVPOQVIHRNVVKLNQFVS
QY 57 SCNYVRGSLVLKLTVYASTFNKGRLMAPFPIMMQGTQRKKHCKLCFLMVCDIGL
Db 426 NATYIFRGSVLRLMSVASTFNKGRLMAGFPFTNTTSEMNAVITCIDGS
QY 117 TPTYTWGNMRPTRGSVGIGMLRIDVLNLRTYNSSSNVAVNCILQVKGNDAKFM
Db 486 TPTYTFSTWRKTNGRPGLGFQEVNLNRLLTYNSSPCPNKVHCIVQGRIGNDAFY

STANDARD; PRT; 2179 AA.

(Rel. 38, Created)
(Rel. 38, Last sequence update)
(Rel. 41, Last annotation update)
protein [Contains: Coat protein VP0 (PIAB); Coat protein
Coat protein VP1 (PID); Core protein 2A; Core protein P2B;
P2C; Core protein P3A; Genome-linked protein VPG (P3B);
P3C (EC 3.4.22.28) (Protease 3C) (P3C); RNA-directed RNA
(EC 2.7.7.48) (P3D)].
3 (strain Williamson) (Human parechovirus 2).
RNA positive-strand viruses, no DNA stage; Picornaviridae;
122962;

DN N.A.
36025; PubMed=9820139;
ighes P.J., Hyypia T., Stanway G.;
analysis of human parechovirus 2 (echovirus 23).";
21. 79:2641-2650(1998).
4: Picornain 3C is a thiol protease that cleaves at certain
as in the polypeptide.
IC ACTIVITY: Selective cleavage of Gln-|-Gly bond in the
cus polypeptide. In other picornavirus reactions Glu may be
ted for Gln, and Ser or Thr for Gly.
IC ACTIVITY: N nucleoside triphosphate = N diphosphate +
1.
: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP0, VP1,
specific enzymatic cleavages in vivo yield mature proteins.
AVAGES ARE CATALYZED BY P3C.
ITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

PROT entry is copyright. It is produced through a collaboration
a Swiss Institute of Bioinformatics and the EMBL outstation -
Bioinformatics Institute. There are no restrictions on its
-profit institutions as long as its content is in no way
this statement is not removed. Usage by and for commercial
quires a license agreement (See <http://www.isb-sib.ch/announce/>)
email to license@isb-sib.ch.
595; CAA06679.1; -
023; -
PRO09003; Cys_Ser_trypsin.
PRO07053; NC.
PRO00605; RNA helicase.
PRO07095; RNA pol DS PS.
PRO01205; RNA pol P3D.
PRO07094; RNA pol_P3vir.
PRO08375; Viral_cap_coat.
10; NC; 1.
10; RNA dep. RNA pol; 1.
0; RNA helicase; 1.
: Coat protein; Core protein; Transferase;
1 RNA polymerase; Hydrolase; Thiol protease.
1 289
COAT PROTEIN VP0.
290 542
COAT PROTEIN VP1.
543 775
COAT PROTEIN P2A.
776 922
CORE PROTEIN P2B.
923 1044
CORE PROTEIN P2C.
1045 1373
CORE PROTEIN P3A.
1374 1490
GENOME-LINKED PROTEIN VPG.
1491 1510
PICORNAIN 3C.
1511 1710
RNA-DIRECTED RNA POLYMERASE.
1711 2179
CELL ATTACHMENT SITE (POTENTIAL).
763 765
PROTEASE 3C (POTENTIAL).
1669 1669
PROTEASE 3C (POTENTIAL).
1687 1687
2179 AA; 245872 MW; 96803C0BB8856664 CRC64;

Query Match 41.8%; Score 413.5; DB 1; Length 2179;
Best Local Similarity 45.9%; Pred. No. 5.4e-34;
Matches 84; Conservative 32; Mismatches 60; Indels 7;

QY 2 DLMEIARMPVYKGERTEPGGTN-----GYFQWSTHSPINWVFDGGHLEDM
DB 361 DLVKISQLFSV-MADSTTPSANHGIDQKGYFKWSANSDPQAIVHRNLVHLNLF
QY 57 SCVNYWRGSLVLTUYASTNKGRLMAFPIMMGTQKQKHKCLFWVCDTGL
DB 420 NSYSYFRGSLIRLSVYASTNKGRLN-GFPFNSSTDETSEIDNAIYITCIDGS
QY 117 TIPYTWGNWPTRGSVIGRLIDVNLNLTYNSSSPNAVNCILOVKGNDKAFW
DB 479 TIPYFSFTWMTKTHGKPIGLFQIEVNLNLTYNSSSPNEVYCIQVQKMGQDAKFF
QY 177 IWV 179
DB 539 VTF 541

RESULT 4
ID POLG_TMEVG STANDARD; PRT; 2303 AA.
AC P08545; Q88593; Q88594;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Coat protein VP1 (PIA); Coat prot
DE (PIB); Coat protein VP3 (PIC); Coat protein P2B; Core protein P2C; Core
DE (EC 3.4.22.29) (P2A); Core protein P2B; Picornain 3C (EC 3.4.22.2
DE P3A); Genome-linked protein VPG (P3B); Picornain 3C (EC 2.7.7.48)
DE (Protease 3C) (P3C); RNA-directed RNA polymerase (EC 2.7.7.48)
OS Theiler's murine encephalomyelitis virus (strain GDVII) (TMEV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornavi
OC Cardiovirus.
OX NCBI_TaxID=12127;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=88265847; PubMed=2838951;
RA Pevear D.C., Borkowski J., Calenoff M., Oh C.K., Ostrowski E.,
RA Lipton H.L.;
RT "Insights into Theiler's virus neurovirulence based on a genom
RT comparison of the neurovirulent GDVII and less virulent BeAn
RT strains."
RL Virology 165:1-12(1988).
RN [2]
SEQUENCE OF 1704-2303 FROM N.A.
RX MEDLINE=87061197; PubMed=3023668;
RA Ozden S., Tangy F., Chamorro M., Brahic M.;
RT "Theiler's virus genome is closely related to that of
RT encephalomyocarditis virus, the prototype cardiovirus."
RL J. Virol. 60:1163-1165(1986).
CC -!- FUNCTION: It is thought that the P2C protein attaches to ve
CC membranes and is associated with viral RNA synthesis.
CC -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CER
CC O/G SITES IN THE POLYPEPTIDE. IT MAY BE A CYSTEINE PROTEASE
CC -!- CATALYTIC ACTIVITY: Selective cleavage of Gln-|-Gly bond in
CC poliovirus polyprotein. In other picornavirus reactions Glu
CC substituted for Gln, and Ser or Thr for Gly.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphospha
CC (RNA) (N).
CC -!- SUBUNIT: The virus capsid is composed of 60 icosahedral unil
CC each of which is composed of one copy each of proteins VP1,
CC VP3, and VP4.
CC -!- PTM: Specific enzymatic cleavages in vivo yield mature prot
CC -!- SIMILARITY: CLOSELY RELATED TO ENCEPHALOMYOCARDITIS VIRUS.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC This SWISS-PROT entry is copyright. It is produced through a co
CC between the Swiss Institute of Bioinformatics and the EMBL o
CC the European Bioinformatics Institute. There are no restricti

non-profit institutions as long as its content is in no way
and this statement is not removed. Usage by and for commercial
requires a license agreement (See <http://www.isb-sib.ch/announce/>
email to license@isb-sib.ch).

```

2; AAA47929.1; -
3; AAA47933.1; -
; A26100.
; 1; 1TME.
; 010; -
[PR0009003; Cys_Ser_trypsin.
[PR001676; Rnv.
[PR000605; RNA helicase.
[PR007095; RNA_pol_DS_PS.
[PR001205; RNA_pol_P3D.
[PR007094; RNA_pol_PSVir.
[PR008975; RNA_cap_coat.
]73; rhv; 3.
]80; RNA_dep_RNA_pol; 1.
]10; RNA_helicase; 1.
; Coat protein; Core protein; Transferase;
; RNA polymerase; Hydrolase; Thiol protease; Myristate;
. 1 76 LEADER PEPTIDE.
77 147 COAT PROTEIN VP4.
148 414 COAT PROTEIN VP2.
415 646 COAT PROTEIN VP3.
647 922 COAT PROTEIN VP1.
923 1064 CORE PROTEIN P2A.
1065 1191 CORE PROTEIN P2B.
1192 1517 CORE PROTEIN P2C.
1518 1605 CORE PROTEIN P3A.
1606 1625 GENOME-LINKED PROTEIN VP3.
1626 1842 PICORNAIN 3C.
1843 2303 RNA-DIRECTED RNA POLYMERASE P3D.
77 N-myristoyl glycine (in host) (By
similarity).
1793 1793 PROTEASE (POTENTIAL).
1811 1811 PROTEASE (POTENTIAL).
1747 1750 VTGI -> CYRD (IN REF. 2).
DDVYQSLF -> GRCLPIIF (IN REF. 2).
2003 2003 E -> Q (IN REF. 2).
2008 2008 E -> H (IN REF. 2).
2046 2046 F -> L (IN REF. 2).
2128 2134 LIRGLP -> YSWGPA (IN REF. 2).
2303 AA; 256342 MW; 5D0FB5E47F72A04 CRC54;
14.98; Score 148; DB 1; Length 2303;
ilarity 25.5%; Fred.No. 9.2e-07;
Conservative 34; Mismatches 87; Indels 16; Gaps 5;
MEIARMPGVYKGERTEPGGTNGYFOWSHTHS-PINWVFDGGIHLDMENLNFSSC-- 58
LELCKLPFLGNPSID---NKRYPVFSATNSVPATSLVDYQVASCSTANSLAAVA 513
YNYWRGSIVLKLTVASTFNKGRURMAFP--IMQGQRKKHKCLFMCVCDIGLNTTF 114
FNQYRGLNLFLEVTGAMVAGKFRATYTPFGAGKPTTRDQAMQATVAIWDGLNSSF 573
[TIP-----YTWGNWMPRTGSRVIGWLRLDVLNLTYNSSSPNAVCLIQVMGNDAK 168
TAPFISPHYRQTSYTSPTITSVDGWVTWQLTPLTPSPGPTTHSDILTLVSAGDDFT 633
]VP 172
;
MP 637
STANDARD; PPT; 2303 AA.
583; Q88584; Q88585; Q88586; Q88587; Q88588; Q88589;
591; Q88592;

```

```

DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Coat protein VP4 (P1A); Coat prot
DE (P1B); Coat protein VP3 (P1C); Coat protein VP1 (P1D); Picorna
DE (EC 3.4.22.29) (P2A); Core protein P2B; Core protein P2C; Core
DE P3A; Genome-linked protein VP3 (P3B); Picornain 3C (EC 3.4.22.
DE (Protease 3C) (P3C); RNA-directed RNA polymerase (EC 2.7.7.48)
OS Theiler's murine encephalomyelitis virus (strain BeAn 8386) (TM
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornav
OC Cardiovirus.
OX NCBI_TaxID=12125;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=87198877; PubMed=3033278;
RA Pevear D.C.; Calenoff M.; Rozhon E.; Lipton H.L.;
RT "Analysis of the complete nucleotide sequence of the picornavi
RT Theiler's murine encephalomyelitis virus indicates that it is
RT related to cardioviruses.";
RL J. Virol. 61:1507-1516(1987).
RN [2]
RN X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
RX MEDLINE=92196127; PubMed=1312722;
RA Luo M., He C., Toth K.S., Zhang C.X., Lipton H.L.;
RT "Three-dimensional structure of Theiler murine encephalomyelit
RT virus (BeAn strain).";
RL Proc. Natl. Acad. Sci. U.S.A. 89:2409-2413(1992).
CC -|- FUNCTION: It is thought that the P2C protein attaches to v
CC membranes and is associated with viral RNA synthesis.
CC -|- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CSE
CC Q/G SITES IN THE POLYPEPTIDE. IT MAY BE A CYSTEINE PROTEASE;
CC -|- CATALYTIC ACTIVITY: Selective cleavage of Gln-[Gly bond in
CC poliovirus polypeptide. In other picornavirus reactions Glu
CC substituted for Gln, and Ser or Thr for Gly.
CC -|- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosph
CC [RNA] (N).
CC -|- SUBUNIT: The virus capsid is composed of 60 icosahedral uni
CC each of which is composed of one copy each of proteins VP1,
CC VP3, and VP4.
CC -|- PTM: Specific enzymatic cleavages in vivo yield mature prot
CC -|- SIMILARITY: CLOSELY RELATED TO ENCEPHALOMYOCARDITIS VIRUS.
CC -|- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a co
CC between the Swiss Institute of Bioinformatics and the EMBL
CC the European Bioinformatics Institute. There are no restricti
CC use by non-profit institutions as long as its content is
CC modified and this statement is not removed. Usage by and for
CC entities requires a license agreement (See http://www.isb-sib.ch
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M16020; AAA47930.1; -
DR MEROPS; 1TME; 31-OCT-93.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001676; Rnv.
DR InterPro; IPR000605; RNA_helicase.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR001205; RNA_pol_P3D.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR InterPro; IPR008975; Viral_cap_coat.
DR Pfam; PF00073; rhv; 3.
DR Pfam; PF00680; RNA_dep_RNA_pol; 1.
DR Pfam; PF00910; RNA_helicase; 1.
KW Polyprotein; Coat protein; Core protein; Transferase;
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease; Myrista
KW 3D-structure; Lipoprotein.
FT PROPEP 1 76 LEADER PEPTIDE.
FT CHAIN 77 147 COAT PROTEIN VP4.
FT CHAIN 148 414 COAT PROTEIN VP2.
FT CHAIN 415 646 COAT PROTEIN VP3.
FT CHAIN 647 922 COAT PROTEIN VP1.
FT CHAIN 923 1064 CORE PROTEIN P2A.

```

FT	STRAND	691	692	
FT	TURN	707	708	
FT	STRAND	713	714	
FT	TURN	760	764	
FT	STRAND	770	780	
FT	STRAND	791	794	
FT	TURN	796	797	
FT	STRAND	820	822	
FT	STRAND	831	835	
FT	STRAND	844	845	
FT	TURN	865	866	
FT	STRAND	871	874	
FT	STRAND	880	892	
SQL	SEQUENCE	2303 AA; 256280 MW; E2C7737DFDBEB786 CRC64;		
	Query Match	13.9%;	Score 138; DB 1; Length 2303;	
	Best Local Similarity	26.1%;	Pred. No. 9.8e-06;	
	Matches	48; Conservative	33; Mismatches 87; Indels 16;	
QY	2	DLMEIARMSVYKGERTEPGGTNGYFOWSHTS-PINWVFDGGIHL--EDMEN-		
Db	457	DLLELCKLTFELGNPT---NNKRPYPFSATNSVPATSMVDYQVALSCSMA		
QY	57	SCNYWRGSGTVLKLTVYASTFNKGLRMAPP--IMMQGTQRKKHKCLFVWCD		
Db	514	RNFNOYRGSLNELFVFTGAAMVKGLFIAYTPPGAGKPTTRDQAMQSTVAIWD		
QY	115	EMTIP-----TWGNWMEPTRGSGVIGMLRIDVLNRLTYNSSSPNAVNCILOV		
Db	574	NFTAPFISHTHYRQISYTSPTTISVDGWVVMKLTFLTPYSPGTPTNSDILTLV		
QY	169	FMVP 172		
Db	634	LRMP 637		

```

RESULT 6
POLG TMEVD
ID POLG TMEVD STANDARD; PRT: 2301 AA
AC P13699; Q88564; Q88565; Q88566; Q88567; Q88568; Q88569; Q88570;
AC Q88571; Q88572; Q88573; Q88574; Q89580;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Genomic polypeptide [Contains: Coat protein VP4 (PIA); Coat prot
DE (PIB); Coat protein VP3 (PIC); Coat protein VP1 (PID); Picornal
DE (EC 3.4.22.29) (P2A); Core protein P2B; Core protein P2C; Core
DE P3A; Protease-linked protein VPG (P3B); Picornain 3C (EC 3.4.22.2
DE (Protease 3C) (P3C); RNA-directed RNA polymerase (EC 2.7.7.48)
DE Theiler's murine encephalomyelitis virus (strain DA) (TMEV)
DE Viruses; ssRNA positive-strand viruses, no DNA stage; Picornavi
DE Cardiovirus.
OC NCBI_TaxID=12126;
OC [1]
RN SEQUENCE FROM N.A.
RN MEDLINE=88206072; PubMed=2834872;
RP Ohara Y., Stein S., Fu J., Stillman L., Klamon L., Roos R.P.;
RT "Molecular cloning and sequence determination of DA strain of
RT Theiler's murine encephalomyelitis viruses.";
RT Virology 164:245-255(1988).
RN [2]
RN X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
RP MEDLINE=92196057; PubMed=1549565;
RP Grant R.A., Filman D.J., Fujinami R.S., Icenogle J.P., Hogle J.J.
RT "Three-dimensional structure of Theiler virus.";
RT Proc. Natl. Acad. Sci. U.S.A. 89:2061-2065(1992).
CC -!- FUNCTION: It is thought that the P2C protein attaches to ve
CC membranes and is associated with viral RNA synthesis.
CC -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CER
CC O/G SITES IN THE POLYPEPTIDE. IT MAY BE A CYSTEINE PROTEASE
CC poliovirus polypeptide. In other picornavirus reactions Glu
CC substituted for Gln, and Ser or Thr for Glv.
CC -!- CATALYTIC ACTIVITY: Selective cleavage of Gln|-Gly bond in
CC

```

IC ACTIVITY: N nucleoside triphosphate = N diphosphate +
). The virus capsid is composed of 60 icosahedral units,
 which is composed of one copy each of proteins VP1, VP2,
 d VP4.
 specific enzymatic cleavages in vivo yield mature proteins.
 IY: CLOSELY RELATED TO ENCEPHALOMYOCARDITIS VIRUS.
 ITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

 PROT entry is copyright. It is produced through a collaboration
 e Swiss Institute of Bioinformatics and the EMBL outstation -
 n Bioinformatics Institute. There are no restrictions on its
 n-profit institutions as long as its content is in no way
 d this statement is not removed. Usage by and for commercial
 quires a license agreement (see <http://www.isb-sib.ch/announce/>
 email to license@isb-sib.ch).

 1; AAA47928.1; --
 31-JAN-94.
 .010; --
 PR009003; Cys_Ser_trypsin.
 PR001676; Rhv.
 PR000605; RNA helicase.
 PR007035; RNA_pol_DS_PS.
 PR001205; RNA_pol_P3D.
 PR007094; RNA_pol_PSVir.
 PR008975; Viral_cap_coat.
 73; rhv; 3.
 80; RNA dep RNA pol; 1.
 10; RNA_helicase; 1.
 ; Coat protein; Core protein; Transferase;
 d RNA polymerase; Hydrolase; Thiol protease; Myristate;
 e; Lipoprotein.
 1 76
 77 147 LEADER PEPTIDE.
 148 414 COAT PROTEIN VP4.
 415 646 COAT PROTEIN VP2.
 647 920 COAT PROTEIN VP3.
 921 1062 COAT PROTEIN VP1.
 1063 1189 CORE PROTEIN P2A.
 1190 1515 CORE PROTEIN P2B.
 1516 1603 CORE PROTEIN P2C.
 1604 1623 CORE PROTEIN P3A.
 1624 1840 GENOME-LINKED PROTEIN VPG.
 1841 2301 PICORNAIN 3C.
 77 77 N-myristoyl glycine (in host) (By
 similarity)
 1791 1791 PROTEASE (POTENTIAL).
 1809 1809 PROTEASE (POTENTIAL).
 96 97
 103 106
 109 110
 162 166
 167 168
 169 173
 179 181
 182 184
 193 194
 200 201
 204 206
 210 217
 219 220
 223 224
 226 231
 232 232
 233 235
 238 240
 241 247
 248 249
 250 262
 267 268
 270 279
 280 280

FT STRAND 286
 FT TURN 287
 FT STRAND 289
 FT TURN 291
 FT STRAND 294
 FT TURN 295
 FT STRAND 297
 FT TURN 298
 FT STRAND 305
 FT TURN 306
 FT STRAND 324
 FT TURN 325
 FT STRAND 326
 FT TURN 327
 FT STRAND 330
 FT TURN 331
 FT STRAND 332
 FT TURN 333
 FT STRAND 334
 FT TURN 335
 FT STRAND 338
 FT TURN 339
 FT STRAND 342
 FT TURN 343
 FT STRAND 345
 FT TURN 346
 FT STRAND 348
 FT TURN 349
 FT STRAND 352
 FT TURN 353
 FT STRAND 356
 FT TURN 357
 FT STRAND 362
 FT TURN 363
 FT STRAND 364
 FT TURN 365
 FT STRAND 367
 FT TURN 368
 FT STRAND 370
 FT TURN 371
 FT STRAND 373
 FT TURN 374
 FT STRAND 376
 FT TURN 377
 FT STRAND 379
 FT TURN 380
 FT STRAND 382
 FT TURN 383
 FT STRAND 385
 FT TURN 386
 FT STRAND 388
 FT TURN 389
 FT STRAND 390
 FT TURN 391
 FT STRAND 393
 FT TURN 394
 FT STRAND 396
 FT TURN 397
 FT STRAND 399
 FT TURN 400
 FT STRAND 402
 FT TURN 403
 FT STRAND 405
 FT TURN 406
 FT STRAND 408
 FT TURN 409
 FT STRAND 411
 FT TURN 412
 FT STRAND 414
 FT TURN 415
 FT STRAND 417
 FT TURN 418
 FT STRAND 420
 FT TURN 421
 FT STRAND 423
 FT TURN 424
 FT STRAND 426
 FT TURN 427
 FT STRAND 429
 FT TURN 430
 FT STRAND 431
 FT TURN 432
 FT STRAND 434
 FT TURN 435
 FT STRAND 438
 FT TURN 439
 FT STRAND 441
 FT TURN 442
 FT STRAND 444
 FT TURN 445
 FT STRAND 447
 FT TURN 448
 FT STRAND 450
 FT TURN 451
 FT STRAND 453
 FT TURN 454
 FT STRAND 456
 FT TURN 457
 FT STRAND 459
 FT TURN 460
 FT STRAND 462
 FT TURN 463
 FT STRAND 465
 FT TURN 466
 FT STRAND 468
 FT TURN 469
 FT STRAND 470
 FT TURN 471
 FT STRAND 473
 FT TURN 474
 FT STRAND 476
 FT TURN 477
 FT STRAND 479
 FT TURN 480
 FT STRAND 482
 FT TURN 483
 FT STRAND 485
 FT TURN 486
 FT STRAND 488
 FT TURN 489
 FT STRAND 491
 FT TURN 492
 FT STRAND 494
 FT TURN 495
 FT STRAND 497
 FT TURN 498
 FT STRAND 500
 FT TURN 501
 FT STRAND 503
 FT TURN 504
 FT STRAND 506
 FT TURN 507
 FT STRAND 509
 FT TURN 510
 FT STRAND 512
 FT TURN 513
 FT STRAND 515
 FT TURN 516
 FT STRAND 518
 FT TURN 519
 FT STRAND 521
 FT TURN 522
 FT STRAND 524
 FT TURN 525
 FT STRAND 527
 FT TURN 528
 FT STRAND 530
 FT TURN 531
 FT STRAND 533
 FT TURN 534
 FT STRAND 536
 FT TURN 537
 FT STRAND 539
 FT TURN 540
 FT STRAND 542
 FT TURN 543
 FT STRAND 545
 FT TURN 546
 FT STRAND 548
 FT TURN 549
 FT STRAND 551
 FT TURN 552
 FT STRAND 554
 FT TURN 555
 FT STRAND 557
 FT TURN 558
 FT STRAND 559
 FT TURN 560
 FT STRAND 562
 FT TURN 563
 FT STRAND 565
 FT TURN 566
 FT STRAND 568
 FT TURN 569
 FT STRAND 571
 FT TURN 572
 FT STRAND 574
 FT TURN 575
 FT STRAND 577
 FT TURN 578
 FT STRAND 580
 FT TURN 581
 FT STRAND 583
 FT TURN 584
 FT STRAND 586
 FT TURN 587
 FT STRAND 589
 FT TURN 590
 FT STRAND 592
 FT TURN 593
 FT STRAND 595
 FT TURN 596
 FT STRAND 598
 FT TURN 599
 FT STRAND 601
 FT TURN 602
 FT STRAND 604
 FT TURN 605
 FT STRAND 607
 FT TURN 608
 FT STRAND 610
 FT TURN 611
 FT STRAND 613
 FT TURN 614
 FT STRAND 616
 FT TURN 617
 FT STRAND 619
 FT TURN 620
 FT STRAND 622
 FT TURN 623
 FT STRAND 625
 FT TURN 626
 FT STRAND 628
 FT TURN 629
 FT STRAND 631
 FT TURN 632
 FT STRAND 634
 FT TURN 635
 FT STRAND 637
 FT TURN 638
 FT STRAND 640
 FT TURN 641
 FT STRAND 643
 FT TURN 644
 FT STRAND 646
 FT TURN 647
 FT STRAND 649
 FT TURN 650
 FT STRAND 652
 FT TURN 653
 FT STRAND 655
 FT TURN 656
 FT STRAND 658
 FT TURN 659
 FT STRAND 661
 FT TURN 662
 FT STRAND 664
 FT TURN 665
 FT STRAND 667
 FT TURN 668
 FT STRAND 670
 FT TURN 671
 FT STRAND 673
 FT TURN 674
 FT STRAND 676
 FT TURN 677
 FT STRAND 679
 FT TURN 680
 FT STRAND 682
 FT TURN 683
 FT STRAND 685
 FT TURN 686
 FT STRAND 688
 FT TURN 689
 FT STRAND 691
 FT TURN 692
 FT STRAND 694
 FT TURN 695
 FT STRAND 697
 FT TURN 698
 FT STRAND 700
 FT TURN 701
 FT STRAND 703
 FT TURN 704
 FT STRAND 706
 FT TURN 707
 FT STRAND 709
 FT TURN 710
 FT STRAND 712
 FT TURN 713
 FT STRAND 715
 FT TURN 716
 FT STRAND 718
 FT TURN 719
 FT STRAND 721
 FT TURN 722
 FT STRAND 724
 FT TURN 725
 FT STRAND 727
 FT TURN 728
 FT STRAND 730
 FT TURN 731
 FT STRAND 733
 FT TURN 734
 FT STRAND 736
 FT TURN 737
 FT STRAND 739
 FT TURN 740
 FT STRAND 742
 FT TURN 743
 FT STRAND 745
 FT TURN 746
 FT STRAND 748
 FT TURN 749
 FT STRAND 751
 FT TURN 752
 FT STRAND 754
 FT TURN 755
 FT STRAND 757
 FT TURN 758
 FT STRAND 760
 FT TURN 761
 FT STRAND 763
 FT TURN 764
 FT STRAND 766
 FT TURN 767
 FT STRAND 769
 FT TURN 770
 FT STRAND 772
 FT TURN 773
 FT STRAND 775
 FT TURN 776
 FT STRAND 778
 FT TURN 779
 FT STRAND 781
 FT TURN 782
 FT STRAND 784
 FT TURN 785
 FT STRAND 787
 FT TURN 788
 FT STRAND 790
 FT TURN 791
 FT STRAND 793
 FT TURN 794
 FT STRAND 796
 FT TURN 797
 FT STRAND 799
 FT TURN 800

```

794      795
806      807
810      812
818      822
824      825
828      833
842      843
847      847
849      849
853      854
859      859
860      860
863      864
869      874
878      892

2301 AA; 256159 MW; 0B6095DF153DBDFD CRC64;

12.8%; Score 127; DB 1; Length 2301;
ilarity 24.9%; Pred.No.0.00013;
Conservative 33; Mismatches 78; Indels 34; Gaps 8;

.MEIAKMPVYKGERTEPGTN-GYFQWSHTS-PINWVFDGGIHLEDPNMLFSSC- 58
LEICKLET-FLG---NPNSNNKRYPYFSATNSVPTTSLVDYQVALS-----CSCM 504
-----YNYWRGSGVLKLTLYASTFNKRLRMAFP-IMWQGTORKKHKCLFMV 105
(S)MLAAVARNFNQYRGLNLFVFTGAAMVKGFLIAYTPPGAGKPTTDDQMQATYAI 564
IGLNNTFEMTIP-----YTMGNMMRPTRGVIGMLRDLVNLRTYNSSSPNAVNCIL 159
LGLNSSEVFVTFAPFISPTHYRQTSYTSATIASVDGVTVMQLTPLYPSGAPVNSDILT 624
KMGNDAKFMV 172
: : : :
SAGDDFTLRMP 637

STANDARD; PRT; 834 AA.

(Rel. 12, Created)
(Rel. 15, Last sequence update)
(Rel. 42, Last annotation update)
protein [Contains: Coat protein VP4 (PIA); Coat protein VP2
protein VP3 (PIC); Coat protein VP1 (PID)] (Fragment).
halomyocarditis virus.
RNA positive-strand viruses, no DNA stage, Picornaviridae;
i12107;

OM N.A., AND X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS) .
94193; PubMed=3026048;
and G., Kamer G., Minor I., Arnold E., Rossmann M.G.,
craba D.G., Duke G.M., Palmenberg A.C.;
structure of Mengo virus at 3.0-A resolution.;
:182-191(1987) .

ALLOGRAPHY (3.0 ANGSTROMS), AND REVISIONS TO 384 AND 602.
39145; PubMed=2156078;
Y S., Rossmann M.G.;
refinement and analysis of Mengo virus.";
1. 211:803-844(1990)
: The virus capsid is composed of 60 icosahedral units,
which is composed of one copy each of proteins VP1, VP2,
3 VP4.
: THE PDB DATA BANK CONTAINS THE 3D-STRUCTURE COORDINATE
EINS VP1, VP2, VP3 AND VP4.
15-JUL-92.
31-JAN-94.
P8001676; Rhv.
P8008975; Viral cap coat.

```

483 488
491 491
497 498
506 507
509 510
512 522
525 526
531 539
541 542
544 548
562 564
572 575
587 588
590 594
598 603
607 608
611 609
618 618
623 623
625 627
633 634
636 638
639 640
644 645
648 648
656 657
666 666
669 673
679 690
698 703
705 706
717 718
729 732
735 736
741 745
754 755
759 759
761 761
765 766
772 772
775 776
781 786
789 790
793 805
834 834
834 AA; 91715 MW; 82AF73A382BED104 CRC64;
12.6%; Score 124.5; DB 1; Length 834;
ilarity 23.7%; Pred. No. 7.2e-05;
Conservative 34; Mismatches 74; Indels 43; Gaps 10;
LMEIARMPVYKGERTEPGGTNGYFQWSHT---HSPINWVFDGGIHLEDMFNLFSS 57
FLEIAQIPT-FIGNKV-PNAV-PYIEASNTAVKTQPLA-VYQVTLSCSLANTFLAAL 423
NY--WRGSTVLKLTVVASTENKGLRMAFPP-----IMMQGTQRKKHCLFMV 105
NFAQYRGLSVYTFVTGTAMMKGKFLIAYTPPGAGKPTSRDQAMQAT-----YAI 475
IGLNTTFEMTIPYTWGNWMPTR-----GSLVIGLRLIDVLRNLTYNSSPNA 154
LGLNSSYSFTVPF-----ISPTFRMVGTDQANTNVDGWTWQLTPLTYPPGCPPTS 530
CILQVKGNDAKFMVP 172
ILTMVSAGKDFSLKMP 548

STANDARD; PRT; 901 AA.

(Rel. 27, Created)

DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Coat protein VP4 (PIA); Coat prot
DE (PIB); Coat protein VP3 (PIC); Coat protein VP1 (PID)] (Fragmer
OS Mingo encephalomyocarditis virus (strain 37A).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornavi
OC Cardiovirus.
OX NCBI_TaxID=31702;
RN [1]_SEQUENCE FROM N.A.
RX MEDLINE=92410611; PubMed=1326807;
RA Mann L.M., Anderson K., Luo M., Bond C.W.;
RT "Molecular and structural basis of hemagglutination in mengovir
RL Virology 190:337-345(1992).
CC -1- SUBUNIT: The virus capsid is composed of 60 icosahedral uni
CC each of which is composed of one copy each of proteins VP1,
CC VP3, and VP4.
CC -1- PTM: Specific enzymatic cleavages in vivo yield mature prot
CC This SWISS-PROT entry is copyright. It is produced through a co
CC between the Swiss Institute of Bioinformatics and the EMBL
CC the European Bioinformatics Institute. There are no restricti
CC use by non-profit institutions as long as its content is
CC modified and this statement is not removed. Usage by and for
CC entities requires a license agreement (See <http://www.isb-sib.ch>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M88547; AAB59755.1; -.
DR PIR; A43379; GNNYV.
DR HSP; P12296; LMEC.
DR InterPro; IPR001676; Rhv.
DR InterPro; IPR008975; Viral_cap_coat.
DR Pfam; PF00073; rhv; 3.
KW Polyprotein; Coat protein; Myristate; Lipoprotein.
FT PROPEP 1 67 LEADER PEPTIDE.
FT CHAIN 68 137 COAT PROTEIN VP4.
FT CHAIN 138 393 COAT PROTEIN VP2.
FT CHAIN 394 624 COAT PROTEIN VP3.
FT CHAIN 625 901 COAT PROTEIN VP1.
FT LIPID 68 68 N-myristoyl glycine (in host) (By
FT similarity).
FT NON_TER 901 901
SQ SEQUENCE 901 AA; 99652 MW; CDAB31205DCB4915 CRC64;

Query Match 12.6%; Score 124.5; DB 1; Length 901;
Best Local Similarity 23.2%; Pred. No. 7.9e-05;
Matches 46; Conservative 35; Mismatches 74; Indels 43;
QY 1 KDLMEIARMPVYKGERTEPGGTNGYFQWSHT---HSPINWVFDGGIHLEDMFN 435
Db KDFLEIAQIPT-FIGNKV-PNAV-PYIEASNTAVKTQPLA-VYQVTLSCSLAN
QY 58 CNYV--WRGSTVLKLTVVASTENKGLRMAFPP-----IMMQGTQRKKH
Db SRNFAQYRGLSVYTFVTGTAMMKGKFLIAYTPPGAGKPTSRDQAMQAT-----
QY 106 CDIGLNTTFEMTIPYTWGNWMPTRGSI-----GWLRLIDVLRNLTYN
Db WDLGLNSSYSFTVPF-----ISPTFRMVGTDVLRNLTNADGWTWQLTPLTY
QY 155 VNCILQVKGNDAKFMVP 172
Db AKILTMVSAGKDFSLKMP 615

RESULT 9

POLG_CRPV
ID POLG_CRPV STANDARD; PRT; 426 AA.
AC P13418;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: RNA-directed RNA polymerase

CC	Q/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE	
CC	-1- CATALYTIC ACTIVITY: Selective cleavage of Gln- -Gly bond in	
CC	poliovirus polyprotein. In other picornavirus reions Gln	
CC	substituted for Gln, and Ser or Thr for Gly.	
CC	-1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosph	
CC	{RNA} (N).	
CC	-1- SUBUNIT: The virus capsid is composed of 60 icosahedral un	
CC	each of which is composed of one copy each of proteins VP1,	
CC	VP3, and VP4.	
CC	-1- PTM: Specific enzymatic cleavages in vivo yield mature prot	
CC	-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a c	
CC	between the Swiss Institute of Bioinformatics and the EMBL	
CC	the European Bioinformatics Institute. There are no restricti	
CC	use by non-profit institutions as long as its content is	
CC	modified and this statement is not removed. Usage by and fo	
CC	entities requires a license agreement (See http://www.isb-sib.ch)	
CC	or send an email to license@isb-sib.ch .	
CC	-----	
DR	EMBL; X00463; CAA25152.1; -.	
DR	PIR; A03906; GNNYE.	
DR	HSSP; P12296; 2MEV.	
DR	MEROPS; C03_009; -.	
DR	InterPro; IPR009003; Cys_Ser_trypsin.	
DR	InterPro; IPR001676; Rhv.	
DR	InterPro; IPR000605; RNA helicase.	
DR	InterPro; IPR007095; RNA pol DS PS.	
DR	InterPro; IPR001205; RNA pol_P3D.	
DR	InterPro; IPR007094; RNA_pol_psvir.	
DR	InterPro; IPR008975; Viral_cap_coat.	
DR	Pfam; PF000073; rhv; 3.	
DR	Pfam; PF00680; RNA dep RNA_pol; 1.	
DR	Pfam; PF00910; RNA helicase; 1.	
KW	Polyprotein; Coat protein; Core protein; Transferase;	
KW	RNA-directed RNA polymerase; Hydrolase; Thiol protease; Myrista	
KW	Lipoprotein.	
FT	PROPEP 1 67 LEADER PEPTIDE.	
FT	CHAIN 68 136 COAT PROTEIN VP4.	
FT	CHAIN 137 391 COAT PROTEIN VP2.	
FT	CHAIN 392 622 COAT PROTEIN VP3.	
FT	CHAIN 623 910 COAT PROTEIN VP1.	
FT	CHAIN 911 1056 PICORNAIN 2A.	
FT	CHAIN 1057 1192 CORE PROTEIN P2B.	
FT	CHAIN 1193 1517 CORE PROTEIN P2C.	
FT	CHAIN 1518 1605 CORE PROTEIN P3A.	
FT	CHAIN 1606 1625 GENOME-LINKED PROTEIN VPG.	
FT	CHAIN 1626 1830 PICORNAIN 3C.	
FT	CHAIN 1831 2290 RNA-DIRECTED RNA POLYMERASE.	
FT	LIPID 68 68 N-myristoyl glycine (in host) (By	
FT	ACT_SITE 1784 1784 PROTEASE (POTENTIAL).	
FT	ACT_SITE 1802 1802 PROTEASE (POTENTIAL).	
FT	SEQUENCE 2290 AA; 255756 MW; 26BC81BB7CF68CB5 CRC64;	
FT	-----	
QY	Query Match 12.4%; Score 122.5; DB 1; Length 2290;	
Db	Best Local Similarity 20.8%; Pred. No. 0.00038;	
QY	Matches 43; Conservative 34; Mismatches 69; Indels 61;	
QY	59 -----YNYWRGSTVLKTVYASTFNKGLRMAFPP-----IM	
Db	480 LANTFLAALSRRFAQIGSLVYTVFTGTAMMKGFLIAYTPPGAGKPTSRDQA	
QY	97 KXHKCLFMVDCIGLNTPEMTIPYTWGNMPTRGSVI-----GWLRI	
Db	538 -----YAIWDLGLNSSYSFIVFP-----ISPTFRWGTQVNIADGHWTV	
QY	146 TYNSSPNAVNCILQVKGNDAKFMVP 172	

FT	CHAIN	1520	1607	CORE PROTEIN P3A.
FT	CHAIN	1608	1627	GENOME-LINKED PROTEIN VPB.
FT	CHAIN	1628	1832	PICORNAIN 3C.
FT	CHAIN	1833	2292	RNA-DIRECTED RNA POLYMERASE.
FT	CHAIN	1833	2292	N-methylated glycine (in host. similarity).
FT	LIPID	68	68	
FT	ACT SITE	1786	1786	PROTEASE (POTENTIAL).
FT	ACT SITE	1804	1804	PROTEASE (POTENTIAL).
FT	ACT SITE	2232	2232	PROTEASE (POTENTIAL).
FT	SEQUENCE	2232	2232	AA; 255495 MW; 8540D0EB1437E8D4 CRC64

FT	ACT_SITE	1786	1786	PROTEASE (POTENTIAL).
FT	ACT_SITE	1804	1804	PROTEASE (POTENTIAL).
SQ	SEQUENCE	2292 AA;	255495 MW; 8540D0EB1437E8D4 CRC64;	

VP2 protein [Contains: Coat protein VP4 (Rho); Coat protein VP2 at protein VP3 (Gamma); Coat protein VP1 (Alpha); Picornain 22.29] (P2A) (G); Core protein P2B (I); Core protein P2C protein P3A; Genome-linked protein VP0 (P3E) (H); Picornain 22.28] (protease 3C) (P3C) [p22]; RNA-directed RNA (EC 2.7.7.48) (P3D) (E1).

RNA positive-strand viruses, no DNA stage; Picornaviridae; Picornavirus (serum EMC 3 nonradiotogenic);

12105;

243189; PubMed=2541543;

differences between the diabetogenic and nondiabetogenic

70:282-287(1989).

TES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.

virus polyprotein. In other picornavirus reactions Glu may be mutated for Gln, and Ser or Thr for Gly.

1) .

which is composed of one copy each of proteins VP1, VP2,

specific enzymatic cleavages in vivo yield mature proteins.

entry is copyright. It is produced through a collaboration

an Bioinformatics Institute. There are no restrictions on

Use by and for commercial purposes is not removed.

 email to license@td-sid.ch .

1; GNNYEB.

1.009; -.

PR001676; RHV.
PR000605: PNA helices

PR001205; RNA pol_P3D.

PR008975; viral cap coat.

580; RNA dep RNA_pol; 1.

id RNA polymerase: Hydrolase: Thiol protease: Muriestato:

1 67 LEADER PEPTIDE.

138
393
COAT PROTEIN VP2.

625	901	COAT PROTEIN VP1.
003	1050	DISCUSSING 2

1059	1194	CORE PROTEIN P2B.
1195	1519	CORE PROTEIN P2C.

[illegible]

Rossmann M.G.;
molecular-replacement phases for the refinement of the
virus 14 structure.";
[logr. A 44:270-282(1988)].
ALLOGRAPHY (3.0 ANGSTROMS).
39144; PubMed=2156077;
Rossmann M.G.;
the structure of a common cold virus, human rhinovirus
at a resolution of 3.0 A.";
L. 211:763-801(1990).
P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
SITES IN THE POLYPEPTIDE. IT MAY BE A CYSTEINE PROTEASE.
IC ACTIVITY: Selective cleavage of Gln-I-Gly bond in the
virus polypeptide. In other picornavirus reactions Glu may be
replaced for Gln, and Ser or Thr for Gly.
IC ACTIVITY: N nucleoside triphosphate = N diphosphate +
P.
The virus capsid is composed of 60 icosahedral units,
which is composed of one copy each of proteins VP1, VP2,
VP3, and VP4.
Specific enzymatic cleavages in vivo yield mature proteins.
TRY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
THE PDB DATA BANK CONTAINS THE 3D-STRUCTURE COORDINATE
IN VP1, VP2, VP3 AND VP4.
NOT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
Bioinformatics Institute. There are no restrictions on its
use by any institution as long as its content is in no way
damaging to the institution. This statement is not removed. Usage by and for commercial
purposes requires a license agreement (See <http://www.isb-sib.ch/announcements>
mail to license@isb-sib.ch).

AAA45756.1; -;

CAA25565.1; -;

AAA45758.1; -;

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

5-OCT-94.

InterPro: IPR000605; RNA helicase.
InterPro: IPR007095; RNA_pol_DS_PS.
InterPro: IPR001205; RNA_pol_P3D.
InterPro: IPR007094; RNA_pol_Fsvir.
InterPro: IPR008975; Viral_cap_coat.
Pfam: PF00548; Cys-Protease-3C; 1.
Pfam: PF02226; Pico_P1A; 1.
Pfam: PF01552; Pico_P2A; 1.
Pfam: PF00073; rhv; 3.
Pfam: PF00680; RNA_dep_RNA_pol; 1.
Pfam: PF00910; RNA helicase; 1.
PRINTS: PR00918; CALICIVIRUSNS.
ProDom: PD001125; Cys-Protease_3C; 1.
ProDom: PD001306; Pico_P2A; 1.
ProDom: PD001274; Pico_P2B; 1.
PolyProtein; Coat protein; Core protein; Transferase;
RNA-directed RNA polymerase; Hydrolase; Thiol protease; Myrista
3D-structure; Lipoprotein.
CHAIN 2 69
COAT PROTEIN VP4 (PIA).
COAT PROTEIN VP2 (PIB).
COAT PROTEIN VP3 (PIC).
COAT PROTEIN VP1 (PID).
CORE PROTEIN P2A.
CORE PROTEIN P2B.
CORE PROTEIN P3A.
CORE PROTEIN P3C.
GENOME-LINKED PROTEIN VPG (P3B).
PICORNAVAIN 3C.
RNA-DIRECTED RNA POLYMERASE P3D.
N-myristoyl glycine (in host) (By
similarity).
PROTEASE (POTENTIAL).
PROTEASE (POTENTIAL).
P -> L (IN REF. 3).
I -> T (IN REF. 3).
P -> H (IN REF. 3).
NP -> KS (IN REF. 3).
EMD -> RVG (IN REF. 3).
C -> R (IN REF. 3).
N -> S (IN REF. 3).
P -> L (IN REF. 3).
G -> E (IN REF. 3).
L -> F (IN REF. 3).
L -> H (IN REF. 2).
I -> T (IN REF. 2 AND 3).
I -> V (IN REF. 2 AND 3).
P -> S (IN REF. 3).
P -> A (IN REF. 3).
ACT SITE 1683
ACT SITE 1697
CONFLICT 368
CONFLICT 459
CONFLICT 722
CONFLICT 726
CONFLICT 729
CONFLICT 913
CONFLICT 942
CONFLICT 962
CONFLICT 982
CONFLICT 1193
CONFLICT 1193
CONFLICT 1220
CONFLICT 1399
CONFLICT 1446
CONFLICT 1739
HELI 36
TURN 50
HELI 51
STRAND 57
TURN 63
STRAND 83
TURN 88
STRAND 90
STRAND 101
HELI 103
HELI 113
STRAND 123
HELI 126
TURN 129
STRAND 133
STRAND 138
TURN 142
STRAND 147
TURN 152
HELI 153
TURN 156
HELI 159
HELI 168
TURN 185

CLEAVAGE BETWEEN VP4 AND VP2 IS AUTOCATALYTIC; VP1/VP2A IS CATALYZED BY P2A; ALL OTHER CLEAVAGES ARE CATALYZED BY P3C.

-1- SIMILARITY: P2A PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

-1- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

This SWISS-PROT entry is copyright. It is produced through a cc between the Swiss Institute of Bioinformatics and the EMBL c European Bioinformatics Institute. There are no restricti use by non-profit institutions as long as its content is modified and this statement is not removed. Usage by and for entities requires a license agreement (See <http://www.isb-sib.c> or send an email to license@isb-sib.ch).

EMBL; AF105342; AAF12719.1; -;
EMBL; AF039205; AAD02132.1; -;
EMBL; AF114384; AAF21972.1; -;
HSP; P11404; ID4M.
MEROPS; C03.011; -;
MEROPS; C03.022; -;
InterPro; IPR003593; AAA ATPase.
InterPro; IPR004004; Calci.pol.hel.
InterPro; IPR009003; Cys Ser.trypsin.
InterPro; IPR000139; Pept_3C.picorn.
InterPro; IPR000081; Peptidase_C3.
InterPro; IPR003138; Pico_P1A.
InterPro; IPR002527; Pico_P2B.
InterPro; IPR001676; Rhv.
InterPro; IPR006005; RNA helicase.
InterPro; IPR007095; RNA_pol_DS.PS.
InterPro; IPR001205; RNA_pol_P3D.
InterPro; IPR007094; RNA_pol_P5vir.
InterPro; IPR008975; Viral_cap.coat.
Pfam; PF00548; Cys-protease-3C; 1.
Pfam; PF02226; Pico_P1A; 1.
Pfam; PF00947; Pico_P2A; 1.
Pfam; PF01552; Pico_P2B; 1.
Pfam; PF00073; rhv; 3.
Pfam; PF00680; RNA dep.RNA.pol; 1.
Pfam; PF00910; RNA helicase; 1.
PRINTS; PR00918; CALICVIRUSNS.
ProDom; PD001125; Cys.protease_3C; 1.
ProDom; PD001306; Pico_P2A; 1.
SMART; SM00382; AAA; 1.
PolyProtein; Coat protein; Core protein; Transferase; Myristate
RNA-directed RNA polymerase; Hydrolase; Thiol protease; Lipopro
FT CHAIN 2 69
FT CHAIN 70 330
FT CHAIN 331 568
FT CHAIN 569 850
FT CHAIN 851 1000
FT CHAIN 1001 1099
FT CHAIN 1100 1428
FT CHAIN 1429 1517
FT CHAIN 1518 1539
FT CHAIN 1540 1722
FT CHAIN 1723 2184
FT LIPID 2 2

ACT_SITE 1686 1686
ACT_SITE 1700 1700
CONFLICT 25 25
CONFLICT 447 447
CONFLICT 562 562
CONFLICT 654 654
CONFLICT 986 986
CONFLICT 1110 1110
CONFLICT 2126 2126
SEQUENCE 2184 AA; 243254 MW; D739B8F9E9B033C8 CRC64;

Query Match 9.8%; Score 97.5; DB 1; Length 2184;
Best Local Similarity 21.8%; Pred. No. 0.13;
Matches 45; Conservative 30; Mismatches 84; Indels 47;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

ein search, using sw model

April 23, 2004, 13:57:19 ; Search time 39 Seconds
(without alignments)
1448.148 Million cell updates/sec

KS-09-147-801D-4
90
KDLMEIARMPVYKGRTEP.....QVKMGNDAKFMVPTTSNIW 179

ILQSUM62
lapop 10.0 , Gapext 0.5

017041 seqs, 315518202 residues

its satisfying chosen parameters: 1017041

ngth: 0

ngth: 2000000000

Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

SPTREMBL_25.*
: sp_archaea:*
: sp_bacteria:*
: sp_fungi:*
: sp_human:*
: sp_invertebrate:*
: sp_mammal:*
: sp_mic:*
: sp_organelle:*
: sp_phase:*
: sp_plant:*
1: sp_rodent:*
2: sp_virus:*
3: sp_vertebrate:*
4: sp_unclassified:*
5: sp_rvirus:*
6: sp_bacteriap:*
7: sp_archaeap:*

s the number of results predicted by chance to have a
er than or equal to the score of the result being printed,
ved by analysis of the total score distribution.

SUMMARIES

Query	Length	DB	ID	Description
91.4	507	12	071023	clethronom
90.8	2256	12	08JV19	Q8jv19 ljungan vir
77.8	2253	12	08JV21	Q8jv21 ljungan vir
77.8	2253	12	08JV20	Q8jv20 ljungan vir
69.3	2254	12	Q8ON16	Q8on16 ljungan vir
45.3	2177	12	Q8BES5	Q8bes5 human parec
15.9	2583	12	Q911X8	Q911x8 equine rhin
15.1	818	12	Q56051	Q56051 rhopalosiph
14.2	922	12	Q88498	Q88498 theiller's e
14.2	922	12	Q88495	Q88495 theiller's e
14.2	2303	12	Q88595	Q88595 theiller's m
14.2	2581	12	Q7T7T5	Q7t7t5 heterosigma
13.9	922	12	Q56165	Q56165 theiller's e
13.9	2307	12	Q801V2	Q801v2 theiller-lik
13.5	930	12	Q02472	Q02472 theiller's e
13.4	901	12	Q36967	Q36967 drosophila

17 132 13.3 882 12 Q88487 t
18 129 13.0 895 12 Q91JX3 c
19 127 12.8 920 12 Q88496 t
20 127 12.8 920 12 Q88497 t
21 124.5 12.6 2293 12 Q83422 m
22 124 12.5 2589 12 Q66776 e
23 122.5 12.4 2292 12 Q66763 e
24 122.5 12.4 2292 12 Q89272 e
25 122.5 12.4 2292 12 Q66765 e
26 122.5 12.4 2292 12 Q66850 e
27 122.5 12.4 2292 12 Q8QV04 e
28 122.5 12.4 2292 12 Q7TF87 e
29 110 11.1 746 12 Q8V0G8 f
30 109 11.0 2890 12 Q8B3M2 d
31 109 11.0 2893 12 Q7TG18 d
32 104.5 10.6 746 12 Q8JVE7 f
33 104 10.5 746 12 Q90754 f
34 103.5 10.5 784 12 Q91B35 e
35 103.5 10.5 784 12 Q91B34 e
36 103.5 10.5 784 12 Q91B39 e
37 103.5 10.5 784 12 Q91B41 e
38 103.5 10.5 784 12 Q91B33 e
39 102.5 10.4 784 12 Q91B38 e
40 102.5 10.4 784 12 Q91B40 e
41 102.5 10.4 784 12 Q91B37 e
42 102.5 10.4 2227 12 Q66775 e
43 102.5 10.4 2248 12 Q66774 e
44 102 10.3 811 12 Q36185 p
45 101.5 10.3 784 12 Q91B42 e

ALIGNMENTS

RESULT 1

071023 PRELIMINARY; PRT; 507 AA.
AC 071023;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Polyprotein (Fragment).
OS Clethronomys glareolus picornavirus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornavi
OX NCBI_TaxID=75201;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ljungan 145SL;
RX MEDLINE=99160759; PubMed=10049824;
RA Niklasson B., Kinnunen L., Hornfeldt B., Horling J., Benemar C.,
RA Olof Hedlund K., Mateskova L., Hyypi T., Winberg G.,
RT "A new picornavirus isolated from bank voles (Clethronomys
RT glareolus).";
RL Virology 255:86-93(1999).
DR EMBL; AF020541; AAC12265.1; -;
DR InterPro; IPR008975; Viral_cap_coat.
FT NON TER 507
SQ SEQUENCE 507 AA; 56274 MW; B9FEB34607D3C7A1 CRC64;

Query Match 91.4%; Score 904.5; DB 12; Length 507;
Best Local Similarity 92.2%; Pred. No. 5.9e-89;
Matches 166; Conservative 4; Mismatches 9; Indels 1;
QY 1 KDLMEIARMPVYKGRTEPGGTNGYFQVSHTHSPINWVFDGGIHLHEDMNLNF
Db 318 KDLMEIARMPVYKGRTEPGGTNGYFQVSHTHSPINWVFDGGIHLHEDMNLNF
QY 61 YMRGSTVLKLTIVASTFNKGRLMAPFPIM-MOGTQKXKHKCLFMVCDIGLNNTF
Db 378 YMRGSTVLKLTIVASTFNKGRLMAPFPIMHDARYTEEAQNAIFMVCDIGLNNTF
QY 120 YTWGNWRPTRGSVIGMLRIDVLNRLTYNSSSPNAVNCILQVKGNDAKFMVPTI

WGNWMPTRGSGVIGWLRIDVLNRLTYNSSSPNAVNCILQVKGNDAKFMVPTTSNIW 497

PRELIMINARY; PRT; 2256 AA.

(TrEMBLrel. 22, Created)
(TrEMBLrel. 22, Last sequence update)
(TrEMBLrel. 25, Last annotation update)

us.
RNA positive-strand viruses, no DNA stage; Picornaviridae;
s.
172314;

OM N.A.

L;
54258; PubMed=11955639;

M., Johansson S.;

ic analysis of Ljungan virus and A-2 plaque virus, new
the Picornaviridae.";
85:61-70(2002).

OM N.A.

L;
53720; PubMed=12163611;

.. Niklasson B., Maizel J., Gorbalenya A.E., Lindberg A.M.;
Analysis of Three Ljungan Virus Isolates Reveals a New,
ot Lineage of the Picornaviridae with a Cluster of two
A Proteins.";
6:8920-8930(2002).

OM N.A.

L;
Niklasson B., Gorbalenya A., Lindberg A.M.;
DEC-2000) to the EMBL/GenBank/DBJ databases.

922; AAM46081.1; "

724; F:RNA helicase activity; IEA.

968; F:RNA-directed RNA polymerase activity; IEA.

350; P:transcription; IEA.

079; P:virial genome replication; IEA.

PR009003; Cys_Ser_tryptsin.

PR00605; RNA_helicase.

PR007095; RNA_pol_DS_PS.

PR001205; RNA_pol_P3D.

PR007094; RNA_pol_Psvir.

PR008975; Viral cap_coat.

80; RNA_dep_RNA_pol; 1.

10; RNA_helicase; 1.

2256 AA; 253507 MW; 61FB9677D217A8B6 CRC64;

ilarity 90.8%; Score 898.5; DB 12; Length 2256;

Conservative 91.7%; Pred. No. 1.6e-87;

Mismatches 4; Indels 10; Gaps 1;

LMEIARMPVYKGRTEPGGTNGYFQWSTHSPINWVFDGGIHLNLPNLFSSCVN 60

LMEIARMPVYKGRTEPGGTNGYFQWSTHSPINWVFDGGIHLNLPNLFSSCVN 382

RGSTVLKTVYASTFNKGRLEMAFPFM-MQGTQKKHKLFLMVCDIGLNTFTMTIP 119

RGSTVLKTVYASTFNKGRLEMAFPFM-MQGTQKKHKLFLMVCDIGLNTFTMTIP 442

WGNWMPTRGSGVIGWLRIDVLNRLTYNSSSPNAVNCILQVKGNDAKFMVPTTSNIW 179

WGNWMPTRGSGVIGWLRIDVLNRLTYNSSSPNAVNCILQVKGNDAKFMVPTTSNIW 502

PRELIMINARY; PRT; 2253 AA.

01-OCT-2002 (TrEMBLrel. 22, Created)
01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Polyprotein.

OS Ljungan virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornavi
OC Parechovirus.

OX NCBI_TaxID=172314;

RN [1]_TaxID=172314;

RP SEQUENCE FROM N.A.

RC STRAIN=87-012;

RX MEDLINE=21954258; PubMed=11955639;

RA Lindberg A.M., Johansson S.;

"Phylogenetic analysis of Ljungan virus and A-2 plaque virus, i
members of the Picornaviridae.";
Virus Res. 85:61-70(2002).

RL [2]

RN SEQUENCE FROM N.A.

RP STRAIN=87-012;

RX MEDLINE=22153720; PubMed=12163611;

RA Johansson S., Niklasson B., Maizel J., Gorbalenya A.E., Lindber
"Molecular Analysis of Three Ljungan Virus Isolates Reveals a
Close-to-Root Lineage of the Picornaviridae with a Cluster of 1

Unrelated 2A Proteins.";

RL J. Virol. 76:8920-8930(2002).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=87-012;

RA Johansson S., Niklasson B., Gorbalenya A., Lindberg A.M.;

Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF327920; AAM46079.1; "

DR GO; GO:0003724; F:RNA helicase activity; IEA.

DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.

DR GO; GO:0006350; P:transcription; IEA.

DR GO; GO:0019079; P:virial genome replication; IEA.

DR InterPro; IPR009003; Cys_Ser_tryptsin.

DR InterPro; IPR00605; RNA_helicase.

DR InterPro; IPR007095; RNA_pol_DS_PS.

DR InterPro; IPR001205; RNA_pol_P3D.

DR InterPro; IPR007094; RNA_pol_Psvir.

DR InterPro; IPR008975; Viral cap_coat.

DR Pfam; PF00680; RNA_dep_RNA_pol; 1.

DR Pfam; PF00910; RNA_helicase; 1.

SQ SEQUENCE 2253 AA; 253466 MW; 3C581A32096FCD41 CRC64;

Query Match 77.8%; Score 770.5; DB 12; Length 2253;

Best Local Similarity 79.0%; Pred. No. 1.1e-73;

Mismatches 14; Conservative 14; Indels 3;

QY 1 KDLMEIARMPVYKGRTEPGGTNGYFQWSTHSPINWVFDGGIHLNLPNLF 3

323 KDLMEIARMPVYKGRTEPGGTNGYFQWSTHSPINWVFDGGIHLNLPNLF 3

QY 61 YWRGTVLKVYASTFNKGRLEMAFPFM-MQGTQKKHKLFLMVCDIGLNTFTMTIP 3

383 YWRGTVLKVYASTFNKGRLEMAFPFM-MQGTQKKHKLFLMVCDIGLNTFTMTIP 3

QY 119 PYTGNWMPTRGSGVIGWLRIDVLNRLTYNSSSPNAVNCILQVKGNDAKFMVPTTSNIW 3

442 PYTGNWMPTRGSGVIGWLRIDVLNRLTYNSSSPNAVNCILQVKGNDAKFMVPTTSNIW 3

QY 179 W 179

502 W 502

PRELIMINARY; PRT; 2253 AA.

Q8JV20

Q8JV20

Q8JV20

Q8JV20

Q8JV20

Q8JV20

Q8JV20

Q8JV20

Q8JV20

Q8JV20

Q8JV20

Q8JV20

Q8JV20

Q8JV20

Q8JV20

Q8JV20

Q8JV20

Q8JV20

Q8JV20

Q8JV20

us.
RNA positive-strand viruses, no DNA stage; Picornaviridae;
S.
172314;
OM N.A.
54258; PubMed:11955639;
M., Johansson S.;
ic analysis of Ljungan virus and A-2 plaque virus, new
the Picornaviridae.";
85:61-70 (2002).
OM N.A.
53720; PubMed:12163611;
Niklasson B., Maizel J., Gorbalenya A.E., Lindberg A.M.;
Analysis of Three Ljungan Virus Isolates Reveals a New,
ot Lineage of the Picornaviridae with a Cluster of Two
A Proteins.";
6:8920-8930 (2002).
OM N.A.
Niklasson B., Gorbalenya A., Lindberg A.M.;
DEC-2000) to the EMBL/GenBank/DBJ databases.
921; AAM46080.1; -
724; F:RNA helicase activity; IEA.
968; F:RNA-directed RNA polymerase activity; IEA.
350; P:transcription; IEA.
079; P:viral genome replication; IEA.
PR009003; Cys Ser trypsin.
PR006005; RNA helicase.
PR007095; RNA_pol_DS_PS.
PR001205; RNA_pol_P3D.
PR007094; RNA_pol_PSVir.
PR008975; Viral_cap_coat.
80; RNA_dep_RNA_pol; 1.
10; RNA helicase; 1.
2253 AA; 253484 MW; 1A6B9B07F325C793 CRC64;
ilarity 77.8%; Score 770.5; DB 12; Length 2253;
Conservative 14; Mismatches 21; Indels 3; Gaps 2;
LMEIARMPVYKGERTEPGGTNGYFQWSHSPINWVFDGGIHLDEMPNLPSCCYN 60
LMEIARMPVYKGERTEPGGTNGYFQWSHSPINWVFDGGIHLDEMPNLPSCCYN 382
RGSTVLKTVYASTFNKGRRLMAFFPIMMQG--TQKKHKCLFMVCDIGLNTFEMTI 118
RGSTVLKTVYASTFNKGRRLMAFFPIMMQG--TQKKHKCLFMVCDIGLNTFEMTI 441
IWGNWMEPTRGSVIGLRLIDVLNRLTYNSSPNNAVCILQVKGNDKAFVPTTSNIV 178
IWGNWMEPTRGSVIGLRLIDVLNRLTYNSSPNNAVCILQVKGNDKAFVPTTSNIV 501
179
502
PRELIMINARY; PRT; 2254 AA.
(TrEMBLrel. 24, Created)
(TrEMBLrel. 24, Last sequence update)
(TrEMBLrel. 25, Last annotation update)
us.
RNA positive-strand viruses, no DNA stage; Picornaviridae;

OC Parechovirus.
OX NCBI_TaxID=172314;
RN [1] _taxid=172314;
RP SEQUENCE FROM N.A.
RC STRAIN=M1146;
RX MEDLINE=22541854; PubMed=12655084;
RA Johansson E.S., Niklasson B., Tesh R.B., Shafren D.R.,
Travassos Da Rosa A.P.A., Lindberg A.M.;
RT "Molecular characterization of M1146, an American isolate of L
virus (LV) reveals the presence of a new LV genotype.";
RL J. Gen. Virol. 84:837-844 (2003).
DR EMBL; AF538689; AA083985.1; -
DR GO; GO:0003724; F:RNA helicase activity; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR006005; RNA helicase.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR001205; RNA_pol_P3D.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR InterPro; IPR008975; Viral_cap_coat.
DR Pfam; PF00680; RNA_dep_RNA_pol; 1.
DR Pfam; PF00910; RNA_helicase; 1.
SQ SEQUENCE 2254 AA; 253072 MW; 17569EE279DE06BD CRC64;
Query Match 69.3%; Score 686.5; DB 12; Length 2254;
Best Local Similarity 67.6%; Pred. No. 1.3e-64;
Matches 123; Conservative 28; Mismatches 26; Indels 5;
QY 1 KDLMEIARMPVYKGERTEPGGTNGYFQWSHSPINWVFDGGIHLDEMPNLMI
DB 323 KDLMDLARMPSCLGERAVPQDTTGYFTWSSNITPLNFIYGVYFEDIPNLNI
QY 61 YRGSTVLKTVYASTFNKGRRLMAFFPIMMQG---QRKKHKCLFMVCDIGLNL
DB 383 YRGSTVLKTVYASTFNKGRRLMAFFPIMMQG---QRKKHKCLFMVCDIGLNL
QY 118 IPYTWGNWMEPTRGSVIGLRLIDVLNRLTYNSSPNNAVCILQVKGNDKAFW;
DB 441 IPYTWGNWMEPTRGSVIGLRLIDVLNRLTYNSSPNNAVCILQVKGNDKAFW;
QY 178 VW 179
DB 501 VW 502
RESULT 6
QBES5
ID QBES5 PRELIMINARY; PRT; 2177 AA.
AC QBES5;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Polyprotein.
OS Human parechovirus 3.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornavi
OC Parechovirus.
OX NCBI_TaxID=195055;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A308/99;
RA Ito M., Yamashita T., Tazuki H., Sakae K., Takeda N.;
RT "Isolation and Identification of a Novel Human Parechovirus.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB084913; BAC23086.1; -
DR GO; GO:0003724; F:RNA helicase activity; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR007053; NC.
DR InterPro; IPR006005; RNA_helicase.

```

PR007095; RNA_pol_DS_PS.
PR001205; RNA_pol_P3D.
PR007094; RNA_pol_PSVir.
PR008975; Viral_cap_coat.
70; NC; 1.
80; RNA dep_RNA_pol; 1.
10; RNA_helicase; 1.
1-289
290 545 VP0.
546 771 VP3.
772 920 VP1.
921 1042 2B.
1043 1371 2C.
1372 1488 3A.
1489 1508 3B.
1509 1708 3C.
1709 2177 3D.
2177 AA; 245821 MW; 786866EACE1D72D1 CRC64;
45.1%; Score 448; DB 12; Length 2177;
ilarity 46.9%; Pred.No.6.7e-39; Indels 6; Gaps 2;
Conservative 39; Mismatches 50;
MEIARMPVYKGERTEPGTN-----GYFQWSHTSPINWVFCGGIHLEDMPNLNF 56
IKIAQLFSV-MSDNTPTSSSGIDKYGFDWAATVAPQNMVHRIVVNLAQFPNLTLEM 421
YNYWRGSTVLKLTIVASTFNKGRIRMAFFPIMOGTQKHKKCLFMVCDIGLNTPEM 116
YSYFRGELLIRLSYASTFNKGRIRMGFFPNCTHTQLELDNALIYTCIDIGSDNSFEL 481
PYTWGNMRPRGSGVIGWLRIDVLRNLTYNSSPNAVCILQVKGNDAKFMVPTTS 175
PYSFTWNRKTHGHQLGFLQVEVLNRLTYNSSPNAVCILQVKGNDAKFMVPTTS 175
PRELIMINARY; PRT; 2593 AA.
(TMREMBLrel. 19, Created)
(TMREMBLrel. 19, Last sequence update)
(TMREMBLrel. 25, Last annotation update)
ovirus 3.
RNA positive-strand viruses, no DNA stage; Picornaviridae;
168014;
QM N.A.
/75;
Ficorilli N., Hartley C.A., Wilcox R.S., Weiss M.,
J.;
nitis B virus: A new serotype.";
MAR-2001) to the EMBL/GenBank/DBJ databases.
253; AAK91591.1; -.
028; C.viral capsid; IEA.
524; F:ATP binding; IEA.
724; F:RNA helicase activity; IEA.
968; F:RNA-directed RNA polymerase activity; IEA.
1380; F:structural molecule activity; IEA.
350; P:transcription; IEA.
079; P:viral genome replication; IEA.
PR004004; Calici_pol_hel.
PR009003; Cys_Ser_trypsin.
PR008739; Peptidase_C28.
PR001676; Rhv.
PR000605; RNA_helicase.
PR007095; RNA_pol_DS_PS.
PR001205; RNA_pol_P3D.
PR007094; RNA_pol_PSVir.
PR008975; Viral_cap_coat.

```

```

DR Pfam: PF05408; Peptidase_C28; 1.
DR Pfam: PF00073; rhv; 2. RNA_pol; 1.
DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
DR Pfam: PF00910; RNA_helicase; 1.
DR PRINTS; PR00918; CALICVIRUSNS.
FT CHAIN 1 218 L PROTEIN.
FT CHAIN 219 288 VP4.
FT CHAIN 289 544 VP2.
FT CHAIN 545 774 VP3.
FT CHAIN 775 1093 VP1.
FT CHAIN 1094 1109 2A.
FT CHAIN 1110 1392 2B.
FT CHAIN 1393 1709 2C.
FT CHAIN 1710 1842 3A.
FT CHAIN 1843 1863 3B.
FT CHAIN 1864 2115 3C.
FT CHAIN 2116 2583 3D.
SQ SEQUENCE 2583 AA; 289217 MW; 10880D4E0ED830BC CRC64;
Query Match 15.9%; Score 157.5; DB 12; Length 2583;
Best Local Similarity 25.6%; Pred.No.1.9e-07;
Matches 52; Conservative 34; Mismatches 64; Indels 53;
QY 2 DLMEIARMPVYKGERTEPGTNGYFQWSHTSPINWVFCGGIHLEDMPNLN--
DB 585 DFLQVAVIPTMAE-----VSVGTN-----HKPIP-SFSVGNLTLEDRELLNTI
QY 59 -----YNYWRGSTVLKLTIVASTFNKGRIRMAFFP-----IN
DB 632 TFRNTYLSALALNYQYRGSCICVDFLTGTAMAQKFKVAVYTPPGREPKTLDEP
QY 97 KKHKCLFMVCDIGLNTPEMTPYTWGNW-----RPRGSGVIGWLRIDVLR
DB 690 -----YALWDIGLNSKFKVPIYISASAYRTNEDDRPSVNVNAGNLQVYQIT
QY 150 SSPNAVCILQVKGNDAKFMVP 172
DB 744 NTPHLSDIVFVSAGKDFSLRFP 766
RESULT 8
OS6051 PRELIMINARY; PRT; 818 AA.
AC OS6051;
DT 01-JUN-1998 (TMREMBLrel. 06, Created)
DT 01-JUN-1998 (TMREMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TMREMBLrel. 25, Last annotation update)
DE Structural polyprotein.
OS Rhopalosiphum padi virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Dicistrov
OC Cripavirus.
OX NCBI_TaxID=66834;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=98201645; PubMed=9527915;
RA Moon J.S., Domier L.L., McCoppin N.K., D'Arcy C.J., Jin H.;
RT "Nucleotide sequence analysis shows that Rhopalosiphum padi vi
RL member of a novel group of insect-infecting RNA viruses.";
RL Virology 243:54-65(1998).
RN (2)
RP SEQUENCE FROM N.A.
RA Domier L.L., Moon J.S., McCoppin N.K., Jin H.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF022937; AAC95510.1; -.
DR PIR; T08823; T08823.
DR InterPro; IPR008975; Viral_cap_coat.
KW Polyprotein.
SQ SEQUENCE 818 AA; 90913 MW; 61CED589DDF070AF CRC64;
Query Match 15.1%; Score 149; DB 12; Length 818;
Best Local Similarity 20.5%; Pred.No.3.8e-07;
Matches 44; Conservative 36; Mismatches 83; Indels 52;

```


	Best Local Similarity	27.2%;	Pred No.	1.3e-05;	
	Matches	52;	Conservative	26;	Mismatches 78; Indels 35;
QY	22	GTNGY---	FQWSTHTSPINWVFDGI----	HLEDPN-----	-LNLFSSCY
Db	2121	GKEGYLTFTTWDPSLIVDGKLMDFGPNMYHCPSTINTTGAYALTPLAYFSQEPFI			
QY	67	VKLTVYASTENKGRLRMAFPIMMOGTQRKKHGCLFWVCDDIGLNNTPBMTPIP			
Db	2181	KYRFVVASAFHRGRLRVVDNV-LYSLDAPFNQNFVSVLDTAEQBDFTVVVP			
QY	126	MRPTR-----GSVI-----	GMLRIDVLNRL-TYNSSSNPVANNCILQ		
Db	2240	LENTRPDGVITYGLSYSSVDEAQGNLAIWLNELSTIKSAENVVYVNVV			
QY	168	KFMVETTSNIV 178	:	:	:
Db	2300	RVMPCAEKIM 2310	:	:	:
RESULT 13					
O56165	ID	O56165	PRELIMINARY;	PRT;	922 AA.
AC	O56165;				
DT	01-JUN-1998	(TREMBLrel. 06, Created)			
DT	01-JUN-1998	(TREMBLrel. 06, Last sequence update)			
DT	01-OCT-2003	(TREMBLrel. 25, Last annotation update)			
DE		Polyprotein precursor (Fragment).			
OS		Theiler's encephalomyelitis virus.			
OC		Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae; Cardiovirus.			
NCBI TaxID=12124;					
SEQUENCE FROM N.A.					
STRAIN=low-pathogenic M2 variant of strain BeAn 8386;					
MEDLINE=98105740; PubMed=9444995;					
Kim B.S.; fauch R.L.; Bahk Y.Y.; Kang J.A., Dal Canto M.C., Ha:					
"A spontaneous low-pathogenic variant of Theiler's virus conta-					
mino acid substitution within the predominant VPI(233-250) T-					
epitope.";					
J. Virol. 72:1020-1027(1998).					
EMBL; AF030574; AAC02657.1; -.					
HSP; O88590; LTMF.					
GO; GO:0013028; C:viral capsid; IEA.					
GO; GO:0005198; F:structural molecule activity; IEA.					
InterPro; IPR001676; Rhv.					
InterPro; IPR008975; Viral_cap_coat.					
Pfam; PF00073; rhv; 3.					
Signal.					
SIGNAL.	1	76	POTENTIAL.		
CHAIN	77	>147	CAPSID PEPTIDE VP4.		
CHAIN	148	>414	CAPSID PEPTIDE VP2.		
CHAIN	415	>646	CAPSID PEPTIDE VP1.		
CHAIN	647	>922	CAPSID PEPTIDE VP3.		
NON TER	922	922			
SEQUENCE	922 AA;	101175 MW;	F8C6C64506CFAB6C CRC64;		
Query Watch		13.9%;	Score 138;	DB 12;	Length 922;
Best Local Similarity		26.1%;	Pred. No.	6.8e-06;	
Matches	48;	Conservative	33;	Mismatches 87;	Indels 16;

ROM N.A.
Culley A.I., Suttle C.A.;
a sequence and characterization of a marine picorna-like
NAV, that infects the toxic bloom-forming alga *Heterosigma*
(JUL-2003) to the EMBL/GenBank/DBJ databases.
7486; AAP97137.1; -.
2581 AA; 286412 MW; 9E379A64D7E12574 CRC64;
1.

